POLITECNICO DI TORINO

Literature overview on the automatization of optical networks with the use of ML, for controlling and restoration.

Thesis submitted for the completion of master's degree

(CCNE)

Communication and computer Networks Engineering

Literature overview on the automatization of optical networks with the use of ML, for controlling and restoration.

Author: WASIM SHAH

Supervisor: VITTORIO CURRI

Abstract

In today's world, the Quantity of Data that can be recovered from optical networks (ON, s) is enormously high and diverse, and it's getting huge day by day. Because of this growth it's really a challenging task for network engineers as well for the networks providers and the reason is that, dealing with such huge traffic is off course not an easy task. This discussion is going on from last few decades to make networks more efficient so that they wouldn't need so much human interaction. At the beginning of this discussion it might sounds a bit strange that a network will work without human interaction but thanks to researchers they make it possible and it is not a statement any more. With the help of SDN and AI there are a lot of network portions out there that are running by its own fully or partially. In near future we will see much more modifications and development in this field, and the reason is that the implementation of Machine learning (ML) in networking sector is a hot topic. Vendors are spending a lot of money and time to get the goal of smart networking. ML is the main tool toward this development because Machine learning is an advance tool which is using to analyze the network traffic and pick the useful data from it and take decision based on it, which make possible to run the network efficiently. So, optical networks need serious attention toward self-running networking. That is the reason almost all vendors are working on it with help of ML.

My thesis is also based on overview of some possible research direction toward this development in Optical networks.

Dedication

To my parents and family
without them this success couldn't be possible.

Acknowledgement

I would like to thank prof. VITTORIO CURRI and IHTISHAM KHAN for their guidance and advices.

Contents

Figures: -- 00

Acronym:

ML: Machine Learning
SOFM: Self Organizing Feature Map
BER: Bit Error Rate
EON: Elastic Optical Network
WDM: Wavelength Division multiplexing
SDN: Software define network
NNs: Neural networks
SVM: Support Virtual Machine
SOFM: Self-Organizing Feature Map
RL: Reinforcement Learning
MDP: Markov decision processes
EDFAs: Erbium Doped Fiber Amplifiers.
SPM: Self-phase modulation.
ABNO: Application based network operation.
ANN: Artificial Neural Network.
API: Application programming interface.
ARIMA: Autoregressive integrated moving average.
BBU: base band unit.
BER: bit error rate.
BPSK: Binary Phase Shift Keying.
BVT: Bandwidth variable transponder.
CBR: Case based reasoning.
CD: Chromatic dispersion.
CDR: Call data Record.
CNN: Convolution non-negative matrix factorization.
CO: Central Office.
DBP: Digital Back propagation.
DC: Data Center.
DP: Dual polarization.
DQPSK: Differential quadratic phase shift keying.
DBP: Digital back propagation.
DC: Data center.

EDFA: Erbium learning machine.
ELM: Extreme learning machine.
EM: Expectation Maximization.
EON: Elastic optical network.
FCM: Fuzzy C-Mean clustering.
FN: False negative.
FP: False Positive.
FTTH: Fiber to the home.
GA: Genetic algorithm.
GF: Gain flatness.
GMM: Gaussian Matrix Model.
GMPLS: Generalized Multi-protocol label switching.
GPON: Gigabit passive optical network.
GPR: Gaussian processes nonlinear regression.
HMM: Hidden Markov model.
IP: internet protocol.
ISI: Inter symbol interference.
LDE: Laser drift estimator.
MABP: Multi arm bandit problem.
MDP: Markove decision processes.
MF: Modulation Format.
ML: Machine learning.
MLP: Multi-layer learning.
MPLS: Multi-protocol label switching.
MTTR: Mean time to repaired.
NF: Noise Figure.
NFDM: Network frequency division multiplexing.
NFT: Nonlinear Fourier Transforms.
NLI: Nonlinear interference.
NMF: Non-negative matrix factorization.
NN: Neural network.
NPDM: Network planner and decision maker.
NRZ: Non-return to zero.
OBS: Optical dual binary.
OFDM: Orthogonal frequency division multiplexing.

CHAPTER: 01

Introduction:

1.1 Background and Motivation

Machine learning is giving the idea that, if we get the right information from any network data traces and we use it efficiently, so we would train the machine and these machines will perform various functions based on the training like, how to treat with an issue when it happened in this network. With the use of some powerful tools and mathematical statistics, ML can make machines capable of doing the jobs which human beings are doing normally. If we talk about our concern which is networks, then machine learning is winning the game because of the qualities like it is making complex jobs automated which is in our interest just because several activities concerning the design and operation which almost depend on machines. The above statement is not just a story or theory any more because machine learning is already functioning in various parts of different networks and they are functioning well. Like, Traffic classification, Instruction detection, Cognitive radios. Etc.

In this thesis I focused the use of ML in different area of networking specially in optical networking where ML is used to automatize the network for both controlling and restoration. The use of ML in optical network is important because optical networks are the provider of almost all physical layer infrastructure. Which got numerous properties like low cost, high capacity and many more. These providers are now moving to introduce a new version of telecom sector to data com which will elastic in supporting the properties required these days, like I mention above the increasing of capacity of networks, low cost etc. And its for sure that an upgraded technology will appear soon in market which will overcome all the weaknesses of present networking and for sure it is not possible without the use of automatization and automatization depends on ML at the end. When we talk about automatization the importance of it is to improve different sections of the networking, like wavelength assignment, traffic grooming and routing. My thesis is the simulation of new cross research between Machine learning and optical networks. It is obvious about ML that its applications can be helpful in cross-layer scenario here the analysis on data took place at physical layer, for example bit error rate makes variation at network layer as an example, routing spectrum and presentation of modulation assignment. A possible question which is arises in optical networks is, that why ML is selected as a method which has been useful and considered for last three decades, and till now it is gaining momentum? If we answer this question from a technical perspective, then I would say the current technical progress in optical communication system is just because of the extraordinary growth particularly in the complication in the optical networks. In networking terms, its because of the continues advancement in the complexity of both data plane and control plane. Elastic Optical Network (EON) concept is arisen as new optical network architecture in the data plane which able to upsurge in the need of bounciness in assigning optical network resources. If we talk in compares network the EON proposes an elastic and incessant allocation of bandwidth. Resource allocation is more complex task just because of the elasticity in it and this is a more stimulating for network engineers.

While in control plan, software define network promises to deliver on request reconfiguration and virtualization. But on the other hand, re-configuration in optical scenario is not easy it is more challenging in terms of re-optimization, setting the power of amplifier, fragmentation of spectrum, that causes addition between the optical presentation VDUs while at work on the apparatus level, and control element like SDN and network orchestrators. A high level of freedom and limitation lead the engineers toward a huge series of trials and even to system as well, whenever we talk about the choosing of best network architecture. Machine learning provided a paradigm shift for the emerging optical networks. machine learning makes able the system to get data from different monitors in different parts of the network and then train the machines for future challenges to tackle it by their self. Some of the application of machine learning in optical network are physical flow, prediction of fault, required capacity in future, detection of instruction. Margin of design. The above mention are not the only applications their many more which are coming in different section of my thesis.

1.2 Thesis Objective: The major research objectives proposed in this thesis are as follows:

- The inspiration behind the employment of ML methods in optical communication networks is the rapid and enormous growth in the complexity of the networks from last few years. In my thesis I classify and define related tool that deals with the prerogative of ML to the optical communication and networking.

- I have explained how with the help of some mathematical tool ML makes machines able of performing autonomously intelligent jobs which normally solve by humans by so much efforts and time.

- Among several networking areas, in this thesis I have focused on the use of ML which help to automate optical networking either for controlling either for restoration.

- I have explained why in optical networking ML is so hot topic of research which is investigating from last few years and which is still on gaining speed and attention.

- I have tried to conclude complete overview about some algorithms that are place as ML. because if we want to go in depth of ML approaches its take too much that no single reader or a single overview can explain it all, Reader must need waste verity of books for reference to get complete idea of ML approaches.

1.3 Organization of thesis:

- In this chapter one I describe basic introduction while the remaining are categorized as follow.

- In chapter two I tried to give a brief overview about Optical Networks which is the core Network use in my thesis. Specifically, I explain elastic optical network and the reason behind is that elastic optical network is under concentration for future work.

- In chapter three I tried to explain Artificial intelligence, deep learning and machine learning detail because it is also important to know for the work in automation of optical networks.

- In chapter number four I tried to give a brief overview in general about the algorithms of Machine learning used in optical networks to automatize it. I also explain the working principles of these algorithms.

- This specific chapter number five is about the machine learning using particularly in terms of physical layer and in terms of Network layer. Because of making a network automatize we must consider both the layers physical and network very keenly.

- Chapter number six is important because this chapter is about the discussion on future direction in the field of optical networks automation with the help of intelligent algorithms.

- This is the last chapter of my thesis in this chapter I give a brief conclusion of my thesis. That is the full overview of my thesis.

CHAPTER: 02

Background of optical networks:

2.1 Elastic Optical Network Description:

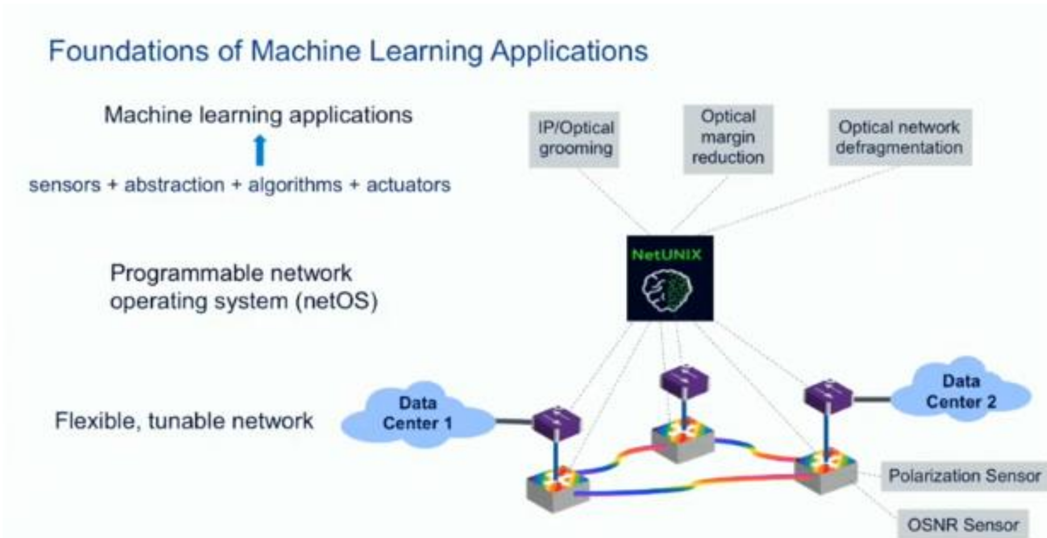


Fig:01. Foundation of ML applications.

Optical network change because of high growth bandwidth-hungry applications such as high-definition video distribution and data-based traffic. Fixed wavelength grid is not fit for 400 G b /s or above and neither fit for more connections. It is noted that the demand for traffic handling has been rising exponentially and rapid development of global connectivity only aggravated the situation. A long-term solution is that to manage how to control the continuous growth in data of the traffic. Spectrum sliced elastic optical networks is lately proposed. With the huge and rapid growth of the communication industry and traffic heterogony another upgraded generation of network is introduced as a solution of the above problem that is EONs which claim to meet communication demand in near future.

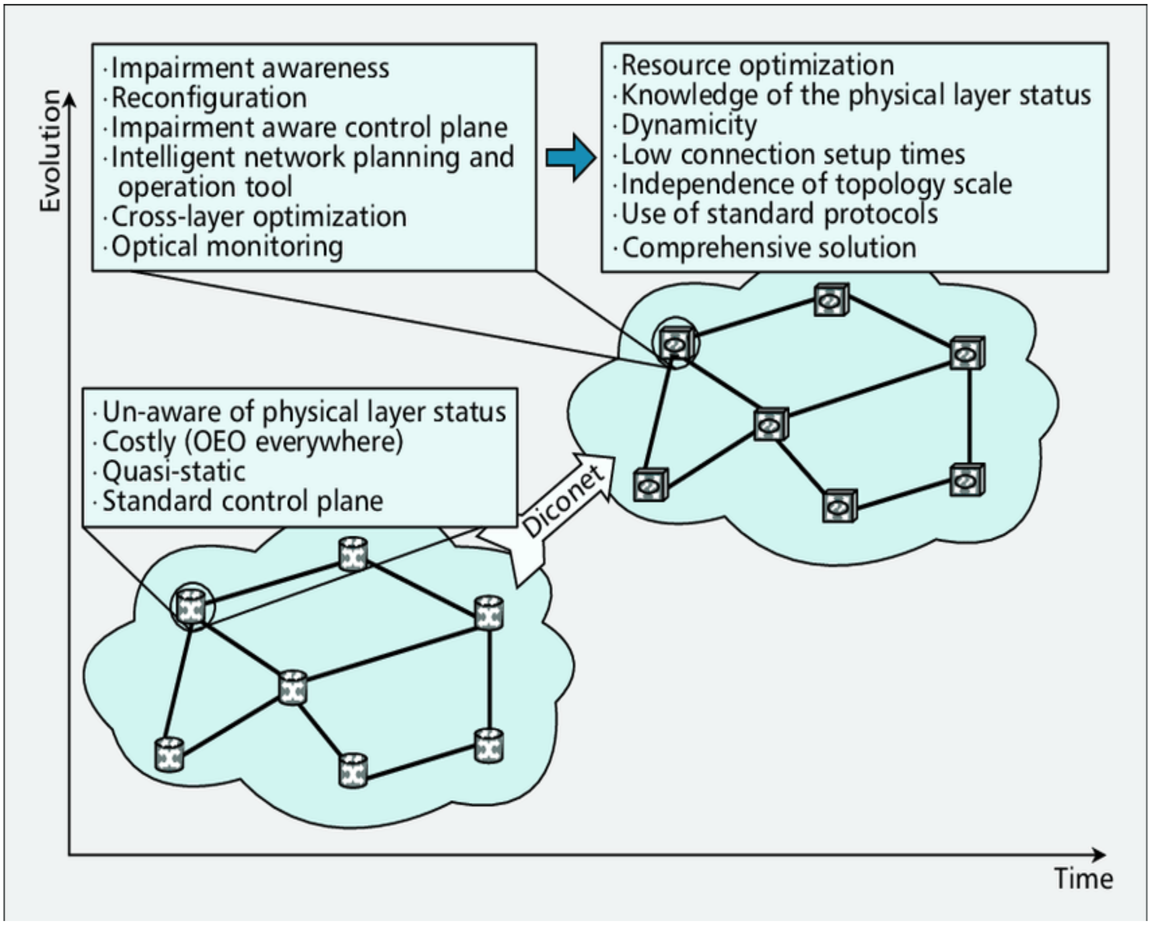


Fig:02. Core Optical Network of Present time:

To encounter upcoming communication requirements. the next age group of long-haul elastic optical networks (EONs) is proposed (driven from dense wavelength division multiplexing (DWDM) networks). Interpretation to the industry standard ITU G.694, 88 channels, 50 GHz to one side, are braced by thick wavelength division multiplexing (DWDM) networks. In the DWDM network, many demands are housed in 50 GHz frequency spaces with somewhat dissimilar center frequencies. The visual range supporting data rate outside 100 Gb/s using standard modulation does not fit in the 50GHz ITU grid due to the predictable use of a fixed grid of 50 GHz between two head-to-head frequency pauses by DWDM. So, DWDM networks are not able to meet the rising needs of communications. Therefore, EONs are recommended to gratify the supplies of the upcoming generation of communications. Conflicting to the conservative DWDM networks, elastic optical networks can use bandwidth variable transceivers (BVT), assemble them appropriate for miscellaneous traffic demands. Fundamentally, elastic optical networks use the continuous elastic optical bandwidth by separating the bandwidth into extremely abundant frequency slots with the enormously small granularity, resulting in the network bandwidth seeming flexible and continuous. Short of the limitation of the 50 GHz ITU grid, elastic optical networks will be able to shift the wider range channels to support high bit rate (such as 400 Gb/s or 1 Tb/s) demands. Thus, elastic optical networks are largely suitable in the future. However, the resources essential to build elastic optical networks (spectrum, regeneration nodes, optical amplifiers, etc.) are incomplete. The RSA exertion of the Planning resource usage of EON has the attention to wide research. The objective of this thesis is to suggest a series of algorithms that can decrease the network resources required to implement EONs. Physical-layer damages of elastic optical networks is considered for the past some years. Valuation of the Physical-layer damages within network stage groundwork. The most widely used model for estimating the Physical-layer damages is the TR. A signal can cover without rebirth. Though, the transmission reach model absences enough elasticity and correctness. This perfect moves toward the lowliest case Physical-layer enhancements in place of considering the real-time network state. When we apply the transmission reach model in unaffected realistic situations of routing and spectrum allocation for EONs, it severely overestimates the Physical-layer impairments. To get a superfluous accurate evaluation of channel Physical-layer impairments, a state dependent model, the Gaussian noise (GN) model, has been suggested. Though, the Gaussian noise model is non-curved and is susceptible from nonlinearities and complication, making it less usable when applied to the RSA problem for elastic optical networks. Therefore, researchers propose an experiential algorithm, the consecutive allocation (SA) algorithm that works at relatively high speed, works with assorted PLI models, and has outstanding performance. The SA algorithm can solve the RSA problem for large network topologies and circulation sizes. In summary, our recommended work uses a linearization of the GN model to guesstimate the PLIs of EONs and resolves the RSA problem over the application of the SA algorithm. Our effort is not only providing a considerable redeemable of resources, but also solves the RSA problem in a reasonably short time. The high scalability as well as the close-to-optimal output of the suggested technique makes it fit for practical networks. [2]

EONs exhibit great potential regarding being highly efficient and flexible, which saves network resources. EONs can support both low transmission rates and high transmission rates simultaneously. EONs can choose a modulation format for each demand that satisfies the QoT requirements through transmission with minimal spectrum usage. However, in conventional DWDM networks, the optical transmission reach, the channel bit rate, and the optical spectrum are fixed. However, some literature considers that full elasticity, i.e., an infinitely small granularity of the sub-carriers, might not be easily accomplished by current techniques. Therefore, less-elastic optical networks, referred to as flexible grid networks, have been proposed as a more realistic version of EONs. Flexible grid networks have a granularity of 12.5 GHz, dividing the spectrum into specific non-overlapping slots. Although the flexibility of the flexible grid network with ‘12.5’ Gigahertz granularities is better than the ITU DWDM with a 50 GHz grid, there is still finite granularity in the network. Through further development of techniques such as more advanced flexible bandwidth transmitters and receivers, the full elasticity of the network can successfully be achieved. In addition, the flexible grid optical network can be considered as a special case of an EON. To make this research more general, this thesis focuses on general EONs instead of flexible grid networks. In summary, there are two main properties of EONs. First, the light-path can be generated with heterogeneous bit rates. Second, the BVT can generate an arbitrary spectrum. These two properties of EONs enable the high efficiency and the flexibility. Because of these properties and the merits of EONs, proper planning for EONs could bring enormous benefits. However, the PLIs are unavoidable in large EONs, especially when we consider that a great number of demands are transmitted in backbone networks. The PLIs affect the channel quality and therefore the quality of the received signal.

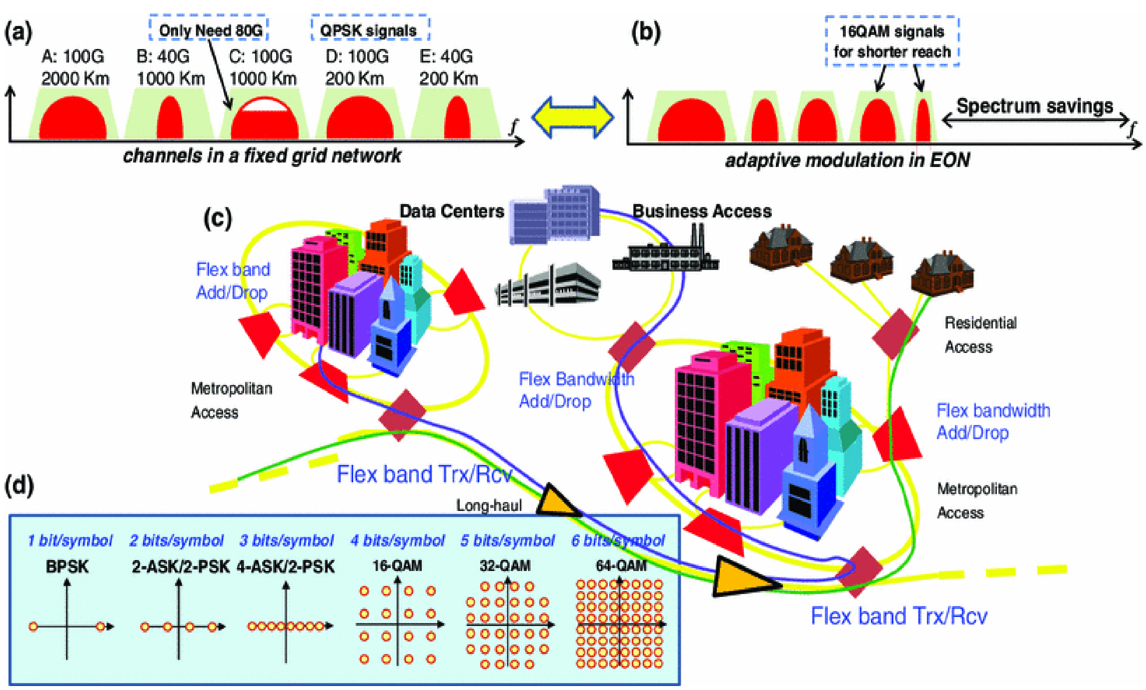


Fig:03. Practical Elastic Optical Network.

2.2 Transmission Reach Model:

As a simpler alternative to the GN model, the TR model is broadly used for estimating PLIs to ensure the QoT is met in long-haul transmission systems. The TR model is applied in greatest investigation lecturing the RSA problematic since of its ease. Additionally, the TR model is linear, so it can easily be implemented in linear programming algorithms. The TR model estimates the longest transparent segment length a signal can travel and still content a traditional approximation of the SINR. Moreover, the parameters of the TR model applied by some researchers are obtained from experimental results. These experimental results are drawn based on different experimental setups, thus lead to questions on the universality of these results. Additionally, the laboratory results are discrete values instead of a continuous function, resulting in model inaccuracies. Instead of implementing the TR model based on experimental data, we tool a GN model founded analytic procedure to generate the parameters of the TR model to make comparison with the 1PM-BPSK: polarization-multiplexed BPSK 2PM-QPSK: polarization-multiplexed QPSK CLGN model fair. In general, because of the state-independence of the TR model, using this model in the network planning stage leads to resource over-provisioning and unnecessary costs. [3]

2.3 Signal Regeneration:

Because the accumulated PLIs constantly harm the systems, the transmitted signal may not satisfy the desired QoT. Consequently, detecting the transmitted signal and recovering the original information may fail at the receiver side. Hence, regeneration nodes that perform optical-electrical-optical (OEO) conversion for reducing the impairments are needed as intermediate nodes. The regeneration (including re-timing, re-shaping and re-amplification) is an electrical process functioning at the intermediate nodes. We assume the PLIs are fully negated through the regeneration process. A plan for allocation of regeneration nodes should account for the high cost of high-speed electronic equipment. This equipment's high cost necessarily implies a similar cost for OEO conversion. These considerations require a careful and conservative of the allocation plan. Because one regeneration circuit can only serve one signal, and a maximal amount of regeneration tours per renewal bulge is expected, not all signals can be regenerated at renewal nodes. And again, the appropriate allocation of regeneration nodes could bring significant benefits. Direction-finding and Range Distribution (RSA) Problem Routing and wavelength allocation (RWA) algorithms are proposed to coordinate the wavelength direction-finding and the assignment simultaneously to obtain the best solution for light-path deployment in fixed grid DWDM networks with 50 GHz frequency spacing. In the conventional RWA problem, direction-finding task for demands are optimized to obtain the minimum resource usage.

The demands in the RSA problem may be deployed with various transmission rate requirements and modulation schemes. In the RWA problem, a demand is transmitted in a 50 GHz frequency slot with a fixed discrete center frequency [33, 35]. However, in EONs, the 50 GHz frequency slot is further divided into infinitely many narrow frequency slots. Therefore, in the RSA problem, a demand is transmitted in a flexible spectrum (a few narrow frequency slots) from its source to its destination. In EONs, without the constraints of a fixed grid in the network, the frequency slots, also known as the spectrum, can be assigned seamlessly. The RSA problem in EONs is to appropriately route the path of the demands and to carefully assign the required spectrum for the demands, to save network resources. Since a demand can be assigned a modulation format that provides desired performance, selection of the modulation formats for each demand along its light-path affects the resources needed by the EONs. Moreover, when regeneration is considered, the noise accumulated along the light path is reduced after the OEO conversion process. Hence, with the implementation of regeneration nodes, constraints based on either the TR or the GN models can guarantee that all demands satisfy the QoT for practical networks. [4]

2.4 Heuristic Method:

Heuristic algorithms are used for solving optimization problems to achieve a tradeoff between the complexity of the problems and a guarantee of optimality. RSA problems are NP-hard [6], usually formulated as MILPs. MILP is an algorithm to realize the best outcome in a mathematical model with linear constraints and objective function. Some variables in MILP are integers, whereas other variables are non-integers [7]. Unlike experiential procedures, MILPs can provide the optimal solution. However, due to the existence of integer variables, which come from the integer decision variables in the RSA problems, MILP solvers must spend a significant amount of time determining the integer variables. Therefore, the optimal solutions are not able to be obtained within a reasonable time using MILPs. Especially with large problem dimensions, obtaining the optimal solutions requires astronomically high computation resources. However, experiential procedures are proposed to solve optimization problems within a reasonable time and obtain near-optimal solutions. Because of the high scalability as well as the less computational resources required, heuristic algorithms have been broadly applied. Accommodates demands in accordance with the length of the routing paths to appropriately coordinate the network resources usage while speeding up the solving process. Proposes an experiential algorithm, mentioned as the R+SA procedure, which decomposes the RSA difficulty into 2 associate difficulties (a direction-finding difficulty plus a spectrum assignment difficulty). After solving the routing problem, the R+SA algorithm then assigns spectrum to these routed light-paths. Heuristic algorithms are efficient sub-optimal algorithms for solving the RSA problem. However, when the complexity of the problem increases, not all variable space is explored within a permitted time, leading to non-ideal performance of these algorithms. [8][9]

CHAPTEER: 03

3.1 A Detailed overview about various Machine Learning Algorithms using in Optical networks:

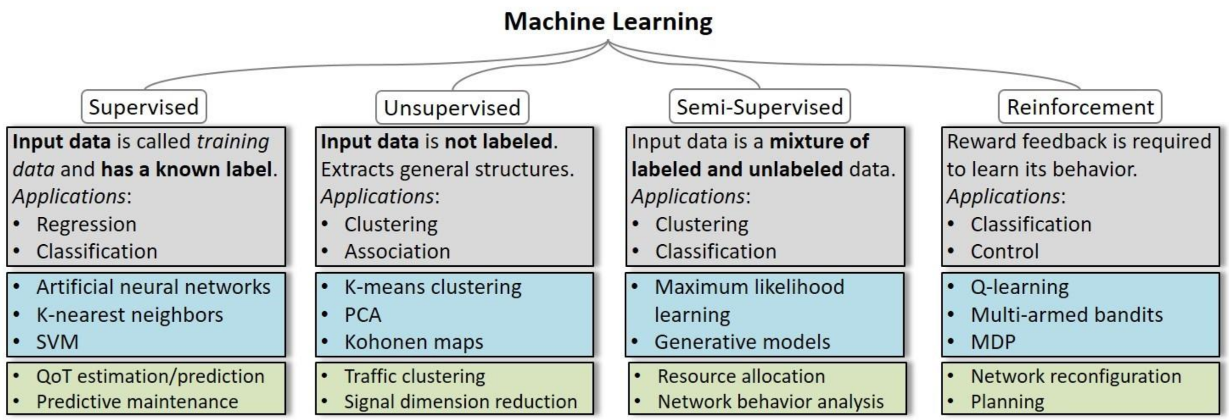


Fig:09. Different machine learning algorithms.

This chapter contain a complete overview about some algorithms that are place as ML. I depth of ML approaches is too much that no single overview can explain it all, Reader must need waste verity of books for reference to get complete idea of ML approaches.
However, I tried to give high level overview about ML approaches. Which is the main and most important part of my overview. Before going to detailed overview, I will present some important basic idea which will be helpful for even first-time reader. To better understand the concept the main idea of ML algorithms I focus on basic three main categories. And these three will be one part of this chapter the second part will be the working principle of these algorithms using in optical networks. [20]

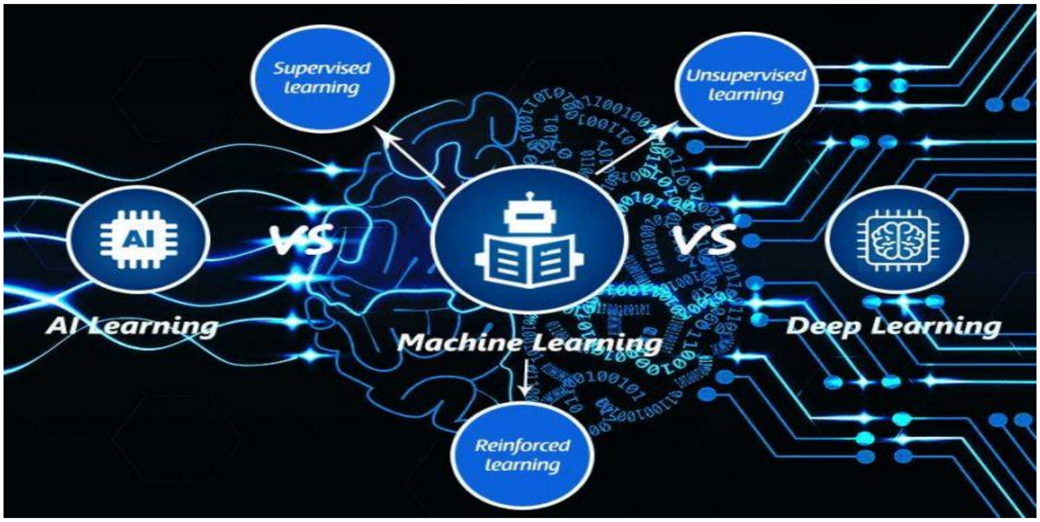


Fig:10. Comparison between AI, ML, and DL.

3.1.1 Supervised learning:
3.1.1.1 PM (Parametric model)
3.1.1.2 Non-PM
3.1.2 unsupervised learning.
3.1.3 Semi-supervised learning.
3.1.4 Re-enforcement learning.
3.1.1 supervised learning:

This is so public to use in the organization issues where the objective to the system capable of learning a classification system which is already existed, in simple words its like a training process for computer. A well known and simple example of classification learning is Digit recognition. Usually, classification learning is good aimed at any problematic wherever making classification is useful and where the classification is easy to control. In this classification problem even, it is not important to give a specific and already arranged classification to each instance of a problem if the agent is solving its own classification. By the way, In organization context this will an example of unsupervised learning.

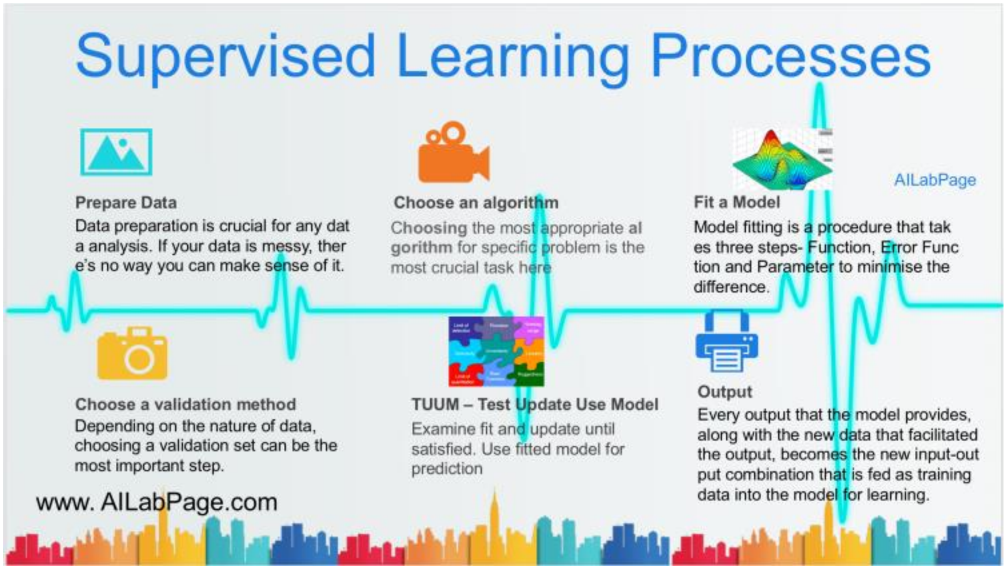


Fig:11. Stepwise Processes in Supervised Learning.

The algorithms work like to create a function which act as a mapping agent between output and inputs. For learning the exact behavior of the function, a standard formulation of supervised learning is needed. The goal of supervised learning is often to make able the computer to learn a classification system that which we already created. Classification learning is efficient where presuming classification is easy to control. Giving pre-determined classification for each instance in a problem is not needed, and when agent work on classification its count as an unsupervised learning. In supervised learning the probability for inputs is usually undefined. This method is not working perfectly if inputs are unavailable. [21] The most common method using for training of the neural networks is supervised learning, in terms of finding the error in a neural network and to make capable the network to minimize the error. On other hand the decision tree instead, the classification in decision tree is used to find which features would provide more information for solving the classification puzzle. In my thesis I focused on both in detail overview. For the time being it is important to keep in mind that both these examples got some supervision in shape of pre-determined classifications. We also need to know that

Inductive is a machine learning process in which a set of rules would learn from instances, for example, a training set, or in more short and simple word it is a classifier which used to take a comprehensive view of new instances from training data set. The implementation of ML in a real-world scenario is showing in the figure 1. In figure first step shows dataset collection, if an expert is reviewing this process he will tell us which field is the most efficient and able for choosing. Let suppose in a case where this method doesn't solve the problem then we got another method called brute force which mean find all in the confidence that correct and pertinent one will be isolated. Even if brute force with dataset together is not suitable for simulation because in most cases it produces noise and absence of feature values. 2nd stage in the information pre-processing and data research which depend on the situation that's why researchers have a huge number of approaches to choose from the lost information. [22] Hodge et al, comes with a review of these days' methods for noise detection. And he also notifies the advantages and disadvantages of this method. This method of instances selection is not only taking over with noise but also tackle the infeasibility of learning form very huge dataset. The selection of instances from these huge datasets is the problem of optimization which efforts to keep the insertion quality while decreasing the sample size. It is also making fast the data mining procedure to work in an efficient way and reduces data from large datasets. We got huge number of procedures for splicing big data sets in samples of instances. Features subset is another subset selection process which point out the unrelated and extra features as much as possible and eliminate them. By doing so this would reduction the measurement of the data and would allow the algorithm to work fast and in an efficient fashion. The fact that each feature is depend somehow on each other mostly effects the efficiency and accuracy of the classification model. But we even got solution for that and solution is to create newly topographies from the rudimentary feature set and researchers called as features transformation. [23] This newly generated feature set may lead a foundation to more accurate classifier. Moreover, the creation of this newly meaningful features can help toward a better free of ambiguities classifier and a healthier understanding of the learn concepts. Hidden Markov models and Bayesian networks uses speech recognition which rely on some components of as well as for finding a fine tune to decreases the quantity of error on inputs. And these inputs are so called training sets. The core purpose of learning algorithms is to minimize the error. For now, I am going to focus on exclusive-or which would demonstrate the combination between one true and one false but never it's come with both false or both true at same time. A usual issue with machine learning algorithm is over fitting of the information and naturally retention the exercise set and then on the bases of this learning an extra simple classification method. But this is also true that not all training sets got the ability to classify the inputs properly. Its mean that if the algorithm is working so perfectly that it even suggests special cases then it would not be appropriate for more general principles, because even this would lead to over fit and this is the actual debate to find an algorithm which works in a moderate fashion, commanding enough to learn composite function and forceful enough to crop generalizable results. [24]

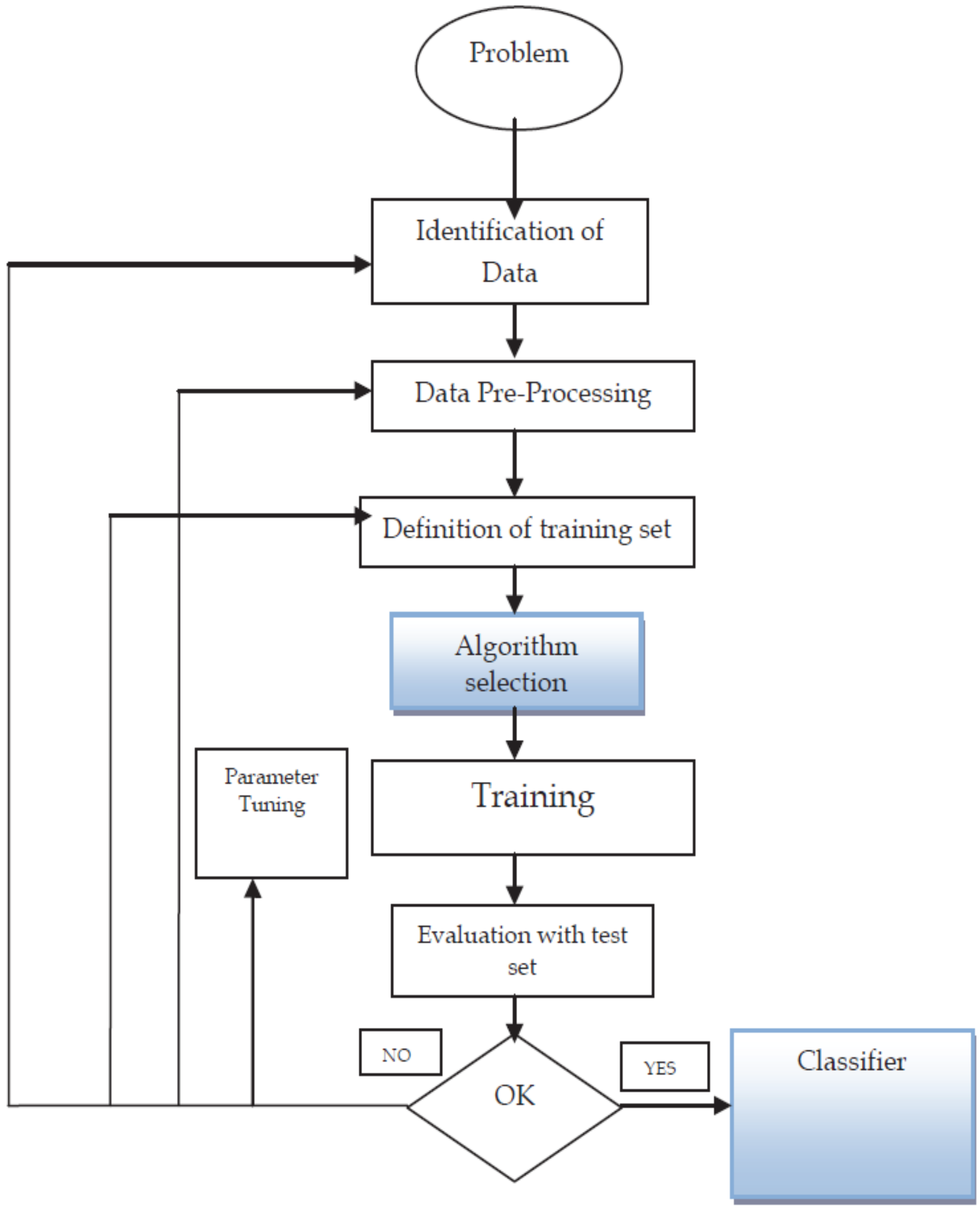


Fig:12. Machine Learning Supervise Process.

Different learning approaches works in a fashion to make a function let say y(x) that lets to guess the worth of the production in the response of a fresh worth of the input. Researches divided oversaw learning in two main categories, and each of them got a huge data to understand it completely I would give a basic touch to both below.

3.1.1.1 Parametric model.

Usually this type of model is using to practice data of training to estimate a parameter W of fixed size. After the completion of the learning stage training data will thrown away while the for further learning processes the parameter w will use only. Linear models are the one which contain a linear mixture of fixed nonlinear basis functions which is for reversion and classification are the modest parametric models in relationships of computation possessions. [25] From polynomial to gaussian to Fourier basis, various ranges are available for basic function. In this case when the number of output values are numerous we got a possibility to use different basis function for each function of all mechanisms. Models like this are liner in terms of limits w, moreover this linearity lead to various advantageous properties.

Yet, their functionality is incomplete with little dimensional contribution interplanetary to the problem. Specific in this section of my thesis I focus on Neural networks (NNs), for now these networks are the efficient once and it's a best example of parametric models. NNs is a combination of units or may say neurons which make a network that's why called NNs. These networks apply an arrangement of functional interchange to the inputs. The very basic function is rummage-sale by every unit in a nonlinear fashion from a linear grouping of the component's input. Respectively neuron got a basic feature that permit any fixed offset in the data. The unfair cases are shared to one usual of parameters by addition a fake input of unitary value to every unit.

Parameters w is assessed in the combination of the constants of linear in the training. Logistic sigmoid and the identity function are the nonlinear functions which are often used. Identity function is the start purpose of output components, while logistic sigmoid purpose for two organization for reversion and multiclass classification problem correspondingly. Distinct influences may cause amid the units from different NNs with diverse features. Units between the input and output of the Neural networks are called hidden units. Network is known as an acyclic graph in the special case of NN. Normally, NN are plane in layers fashion which comprise units in every layer the input is only acceptable from the units which are promptly next and forwarding the output to the exact next layer. [26]

By reducing error role with admiration to the set of limits w a NN is skilled from a given training set. Unlike error functions are used reliant with the sort of problem and the consistent choice of initiation function. Normally, sum of square error is used in case of regression models, on other hand for categorization cross entropy error purpose is used. It is important to notify that error function is non-convex function of the network bounds, and there are many solutions exist for huge ideals.

On the base of incline info numerical methods consider as most common approaches to get the course w which would minimize the mistake function. Error back propagation is an algorithm which is used for NN and which provide a well-organized technique for assessing the offshoots of the mistake function against to w dataset is the one which use the most. Before the training of network, the training set is generally already processed by smearing a lined change and rescale every input variable separately in case of incessant data or discreet ordinal data. And these distorted variables have its unit standard deviation and 0 mean. The same exact method is adopted to the objective standards in case of reversion glitches. On the other hand, in case of discreate categorial data A1 of the coding k structure is rummage-sale. such preprocessing is recognized as feature standardization in addition this is use beforehand the training, till now ML methods are designing to work in fashion that assume all features have its comparable scales.

3.1.1.2 Non-parametric model:

In these methods parameter's numbers are be contingent on the exercise set. These approaches save maybe a subsection, or the total of the exercise set while usage of them among forecast. One of the greatest in use method is k-nearest neighbor models which support vector machines. Two of them will use for regression and cataloguing problems. All training data units are saved in the case of k-nearest approaches. K-nearest examples are retrieved to the input new values during prediction. A voting mechanism is used for classification problem while on other hand for regression problem a voting mechanism is selected to use in which the k nearest samples can be get from predicting mean or median. Iterating with all examples to calculate the neighboring k depending on the dimension is not feasible. For finding k-nearest neighbor in this circumstance, k-trees table will be used. The rudimentary function placed on training samples in SVM, the basic function would be used for the selection of subsets. The amount of training samples and basic functions which would be stowed is generally much slighter then the exercise datasets. SVMs generated a boundary of lined choice with wider probable coldness from the exercise set. Support vectors are stored the once which are closed to the separators. By finding the limits of SVMs a nonlinear optimization issue beside with convex objective function must be solved only for those for whom useful algorithms are present. One of the useful features of SVMs is that if we apply kernel function they can implant data obsessed by a higher space where information points container be linear unglued. Seed purpose job is to measure the similarities between points from the contribution interplanetary. It is articulated the innermost creation form the contribution opinions drew into a higher measurement interplanetary where data should become linearly separated. One of the simplest examples is lined seed its identity function is mapping. Though, if we can present everything based on kernel evaluation, the important thing is to compute the mapping function explicitly in the feature space. One of the most generally used kernel function is gaussian kernel in which the feature space has immeasurable sizes.

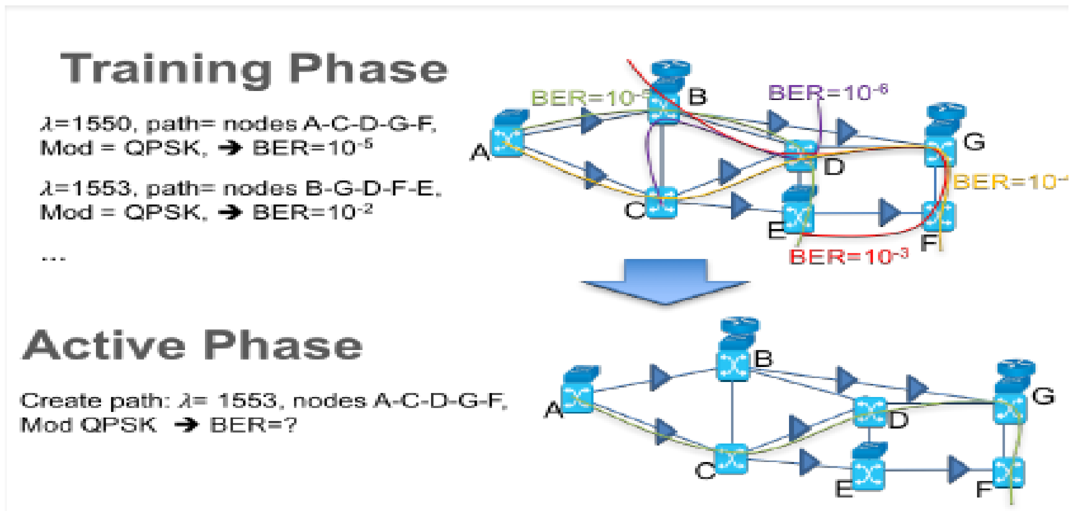


Fig:13. Supervised learning: Algorithm is learned based on dataset that contain paths, modulation, Bit error rate, for new inputs.

3.1.2 Unsupervised learning:

This learning approach is looking much harder the supervised one the objective is to make system able of learning how to perform task which didn't tell by anyone to do and how to do. Unsupervised learning consists of two approaches. The 1st method is to make understand the system not commanding it straightforward categories but instead using some sort of compensation system to identify successful attempts. It is important to note that this kind of practice will usually appropriate in the context of decision about a problem, objective here is to made selection which make the best use of reward.

Method mention above take today world to a very broad view very nicely, Because of the idea of agents reward for doing certain useful actions and punished for taking certain bad actions. Another procedure known as reinforcement learning is used very often aimed at unsupervised learning in where the negotiator or agent job is depending on the previous reward he gets from his performance and punishments. This information is not necessary just since by knowledge and following a prize purpose a go-between can only know what to do in a situation when there is no processing because this agent know only a reward which it learned from somewhere and it would think that is reward is for every sort of action to take. Approach like this will be useful when the calculating each possibility is time taking. [25]

On other hand, this would be too much time taking to learn the trail and error. Never the less, this approach will be too strong because it approves no before discovering classification of examples in other words our classification may not be the best one possible.

A prominent example is predictable perception around game of backgammon were rotate its skull once a series in the computer programs which learned over unsupervised learning method suited tougher form the best then chess player just through playing themselves again and again. Such programs outcome with few ideas that astonished the backgammon specialists and play well than backgammon programs of those whose skilled taking place by pre-classified examples. One more kind of unsupervised learning called clustering its objective is not adventure a purpose, but to discover similarities in the trained data. Belief often is that bunches showing competition sensibly good through an instinctive organization. Till now, gathering is hinge on illustrations which maybe result in a wealthy for clustering in one set while might be poor in another. But algorithm do not have titles for clusters to allocate to them. Algorithm can just produce them and just use those clusters different examples to different clusters. This method is a data determined method which works well when we got abundant data. these days, filtering algorithms use for social platforms for an example amazaone.com declared records are founded on the basic principles of detection alike clusters of individuals and then shuffle the fresh users with the clusters. Sometimes, like in filtering of social traffic the information of any user of a group can be enough for the system to give efficient result. While in some other circumstances it maybe the scenario that the groups are only clue for a specialist who know how to get information from it. Unfortunately, the over smartness of the training data may cause problem even in unsupervised learning. And we got nothing so powerful tool to avoid these problems, the reason is any method who learn from its input would never be strong enough. Unsupervised learning is taken to get structure from the datasets of samples. And the measurement of the excellence of the structure is taken by price purpose which is normally decreasing to deliver finest limits which would categorized the hidden structure from the given collected datasets. As an example, the same structure will never appear in the upcoming groups from the same data sets. If there is no robust ness its recognized as over appropriate in the mechanism learn terminology. In my overview I described the concept of over fitting idea for a group of a gathering model which get a key role in the recovery of info in mainframe requests. There are many top-class successes produced by unsupervised learning, like world champion level backgammon programs, in addition machines having the ability of driving cars.

It is a powerful approach if we have a simple method for assigning value to and action. If we have enough data to form a cluster then clustering is a useful tool, particularly if we have extra data related to other participants of cluster this data can be used to provide additional outcomes because of the dependences in between the data. Classification learning is dominant approach in case when the classifications are well known to be accurate or else when the classifications are basically random elements/events which we want the workstation to be capable of recognizing it for us. A classification learning needed frequently after results get by process would be obligatory as input of something else. Else, it will not be

easy to figure out that input if someone wanted to know what its mean. Both two approaches can be valued and the selection of one will be contingent on the situations like what type of problem is to be solved, time requirement of solving it, clustering and supervised learning is usually quicker than reinforcement learning approach.
If we talk about different applications of different learning algorithms specially an unsupervised learning, they are market research, genius clustering is the successful of all unsupervised learning methods. [26] [27]

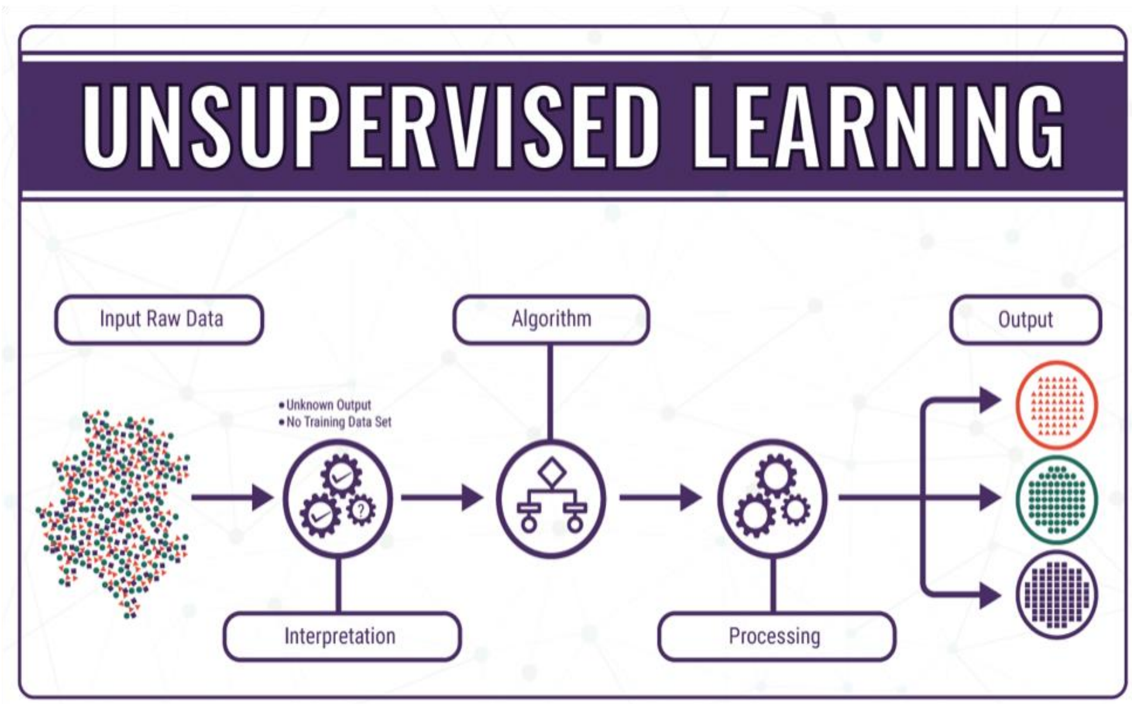


Fig:14. B. Un-SL: notices uncommon shape, in the information sets containing the wavelengths, modulation and bit error rate.

3.1.3 Partial Supervised Learning or semi-supervised.

This learning approaches is got from the combination of previous two supervised and un-supervised presented in above section, this approach address toward problems where majority samples of the

training are unlabeled, even though only limited data points with label are available.
Advantage of this is that in various areas a huge amount of untitled data sets is willingly obtainable. Applications where semi supervised learning which are nearly like the supervised learning.
This type of learning is mostly helpful when the points we have are not too common or so exclusive to get then using of that unlabeled available data points can rise the performance. There are so many old fashions semi-supervised learning, Self-training is one form of it. It’s an iterative process, in first phase only labeled points form the data are used by a supervised learning algorithm. [28] The next step is that at every step few untitled points are labeled rendering the guesses which is resulting for skilled decisions function and these points are then used beside in addition with the unique branded data to rehabilitate by the use of the same oversaw knowledge method. The whole above procedure is mention in figure is mention in fig:4 the idea of the use of labeled and unlabeled data is used in numerous semi supervised learning algorithm after the introduction of self-training.

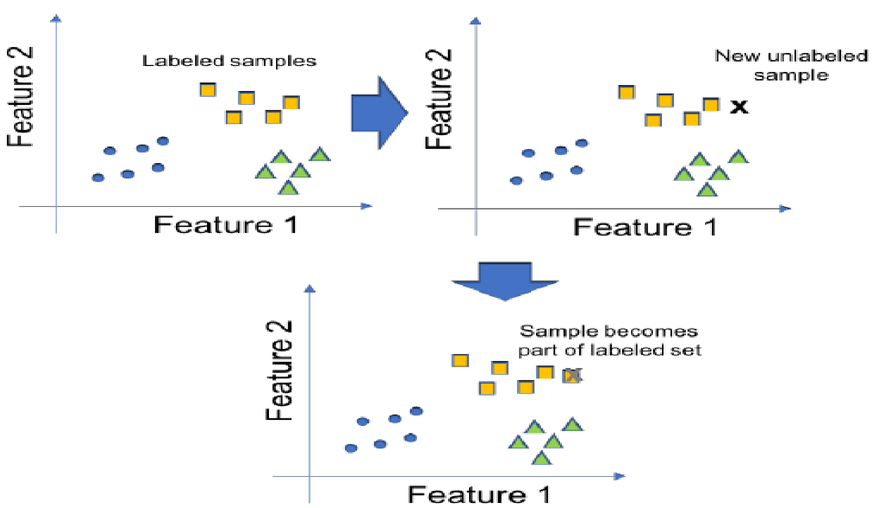


Fig:15. step wise self-training method. Which shows how an unlabeled point is matched with respect to labeled and becoming part of labeled data set.

3.1.4 Reinforcement learning algorithm:

This is also one of the machine learning method in use these days, in many areas its applications are in finance, robotics, etc. final objective of this algorithm is if this algorithm is to skilled itself with a policy. As an example, to play a role of a mapping between in state environment and an action intended to be learn in same time to interact with the environment. Reinforcement learning give permit to agent to skilled from the environment and learn from available actions and make efficient its actions with the use of evaluative reaction. The final goal of the agent is to improve the performance life time. That's why, agents not only consider their rewards but also, they keep in mind the forfeit for the development is future in their behavior Markov decision process is the background of the performance of RL. [29] Awareness of the agent is represented on time k is a state St ϵ *S* in which S is set of environment state. contact of an agent and its environment is depend on its action's performance. Agent performed an action at ϵ A at time k, in which A represent the set of actions belonging to agent, which would cause a new set of states. When a transition happens, an agent can get reward for that, based on the function of reward.

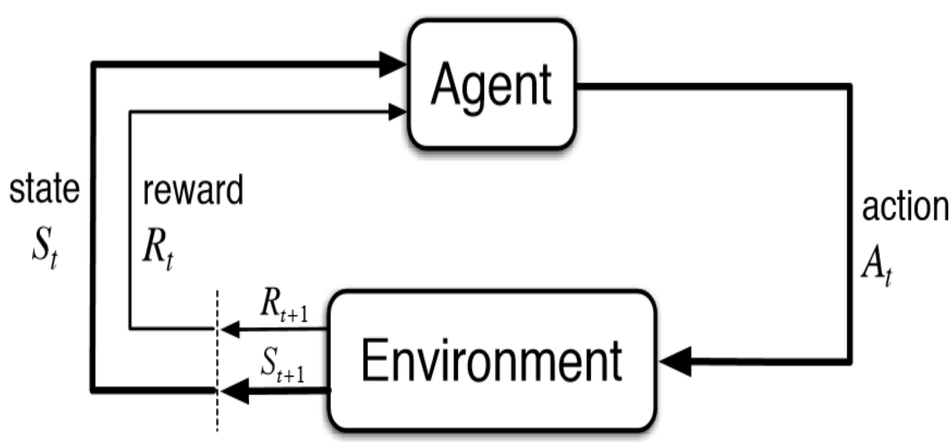


Fig:16. Reinforcement learning steps.

The agent objective is to determine the order of state action pairs which increases the predictability of the discounted award for an example the optimal policy. In the bases of Markov decision processes, this is a proven thing that an ideal inactive deterministic policy exists. We got many algorithms which can learn the best policy in a situation where we got the reward function and states of transitions are quite known and it's called as Model based learning while also when transition states and reward functions stay unknown and this type is known as model free leering. [30]
This the most using reinforcement learning system and its known as Q-learning which is a model free procedure which approximate an ideal value of action. Q is value if action function which is projected reappearance from a pair of state actions for an assumed strategy.

3.1.5 underfitting, overfitting and model section:

In my thesis this section contains the discussion of a familiar issues of ML algorithms also having its solutions. The discussion will focus on both supervised and unsupervised methods. If I would say that underfitting and overfitting are the different flanks of the signal coin, then it is not be wrong. When the model we are using is too complicated for the accessible dataset then overfitting happens. Model in this case will fit the data of training too close, including even the loud examples also outliers then it will result a very bad simplification, like it will predict wrong for new data points. While underfitting is make happen in result of selecting the models that are not too complicated to trace out the important features in data. Example of underfitting, is the usage of linear model to fit quadratic data.
Below fig:17 present alteration among overfitting and underfitting, comparing with a correct model. The measurement of error on measuring data set is just a bad indication on simplification, for the judgement of the performance of the model obtainable set of data split in different two sets first the set of training and the test set. The

training set will use to train trained the model. Another possible option is cross validation which is also very useful when we got limited data set to exploit the training data set as much possible, set of data is divided in k different subsets.

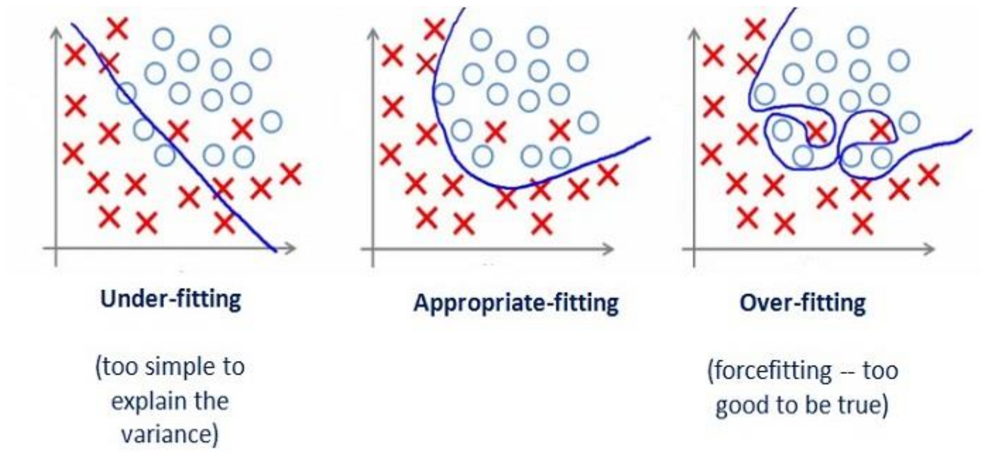


Fig:17. Comparison between under-fitting, appropriate fitting and over-fitting.

By using k subset, the model is trained k times to validation and the reaming k – 1subsets are used for training. With k time runs the performance is average. In overfitting, with test sets the error measurement is higher while it's small in the case of training set. If we talk about underfitting, in both cases test set and the training set error measured is higher. We got many different methods to choose a model which is not exhibition underfitting and overfitting. One of the Possibilities to take a variety of replicas and train it, then compare their performance in a dataset which would be independent, and finally choose one from them having best performance. But still the most common technique is regulation. Which comprise the addition of an additional period, regulation period to the error.

Nevertheless, most well-known method is *regularization*. This method contains the step to add an additional term known as regulation term for the error function usually use among the stages to be train. The regulation period which are simple of altogether is get from the square of all form of the term of regularization are the sum of the square of overall the factors and better-known as decay in weight which drives parameters to zero. Sum of absolute terms of all the factors is one of another common choice. Regularization coefficient is an additional parameter λ, which showing the weights the comparative position of the regularization and error dependent on data sets. If we talk about Neural Networks which having large number of unseen components, present a technique that work in a fashion of arbitrarily eliminating components and its contacts in learning process is exposed to outdo additional regularization approaches.

3.2 Working Principles of Machine learning Algorithms:

Supervised learning is mostly deal with classification. There are different algorithms types which work as supervised learning:

3.2.1 Linear classifier.
3.2.1.1: support vector machine.
3.2.2 Neural networks.

3.2.1 linear classifiers.

Goals of the organization in machine learning are to gather things which having same feature values into different sets. Timothy Jason Shepard present an idea which said that classifier of type linear attains its working by taking a cataloguing choice based on the combination of features in a linear classifier.

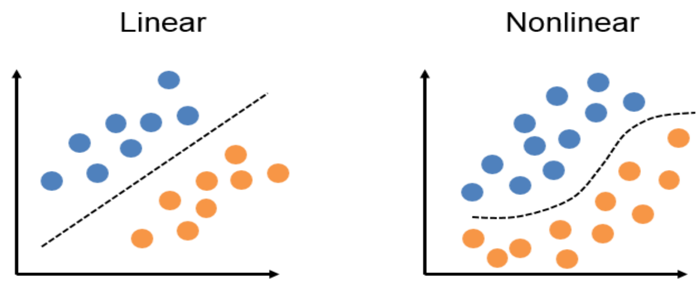


Fig:18. Perception of Linear classifier:

If $\vec{x}$ is the feature vector of the classifier so the output notch is

$$y = f(\vec{w} \cdot \vec{x}) = f\left(\sum_j w_j x_j\right)$$

And $\vec{w}$ is vectors of weight f which is work like a function which conclude the needed outcome from dot product of two vectors. Weightiness of vectors are learned after a training sample set which is titled. Most of the time f represent a sample function plans and the value beyond from a specific limit to the first class while the other values that remains to be placed as second class. [31][32] An extra complex f might gave the chance to an entry fit in to a specific session. Hyper plane is classified as two classes, on one side all points are count as one class known as "yes", while the others are count as "no" classified as second class. The use of a linear classifier is most of the time in such situations when the problem is the speed of classifier next it is too obvious famous as fast classifier.

3.2.1.1 Support Vector Machine: (SVM)

This method is implements as the classification by building n different Dimensions which efficiently splits data into two different classes. Support Virtual Machine models are thoroughly correlated to neural networks. It is obvious, that the use of SVM methodology with a kernel function known as sigmoid which is like two layers and this type of network is called perception neural network.

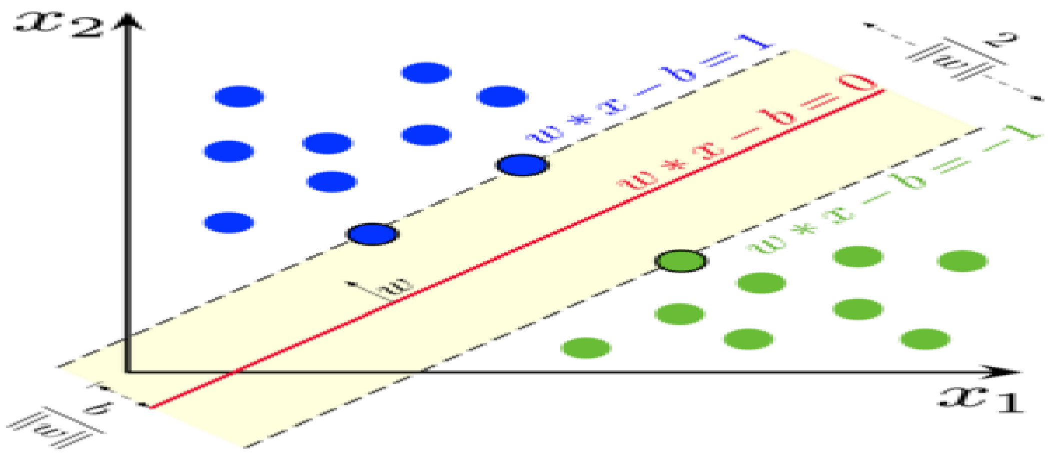


Fig:19. Support Vector Machine Classifier:

SVM methodology is quite like normal perceptron neural network. Just using of a function of SVM from the kernel is another option for exercise technique for multilayer classifiers and the polynomial where heaviness of the network is developed just by the solution of a quadratic problem beside with linear constraint. [33] From phraseology of SVM works a translator which is basically a variable is so-called feature. While an attribute when use for the definition of overexcited flat is called a feature. Task of selecting most appropriate definition is known as collection of features. And a class of features when combined and when it defines a single case it’s called vector. In short, the purpose of SVM is to find out the best of all flat which help in dividing the clusters of vectors in a fashion that all of them lead to one class which is the targeted vector which fall close to the hyper plane. Blow picture presents a view about the support virtual machine process.

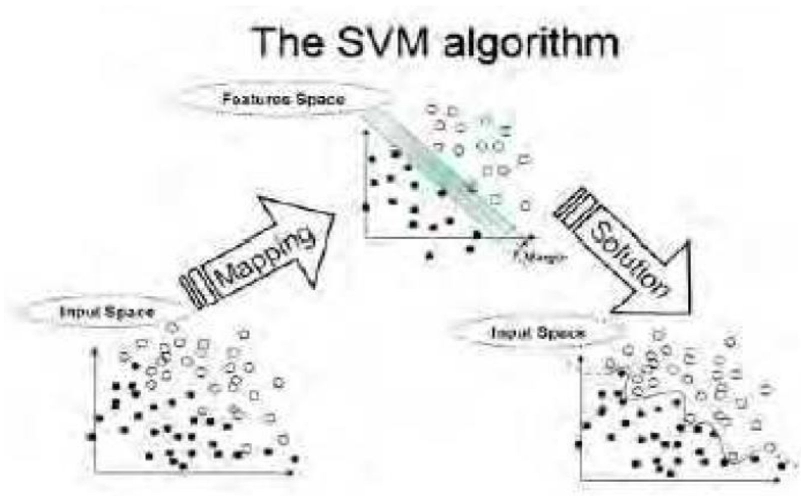


Fig:20. Support Virtual Machine process.

An example with Two-Dimensions only:

Let’s first talk about a simple example of two dimensions, going toward N dimensions hyper planes first we will wish to get a classification and we have two categories for our target variable. Consider it also that we got 2 interpreter variables by means of unceasing standards. If we use these standards for one interpreter with X, and other with Y axis, then the plot we get is might be the picture given down. First class of variables are exemplified with rectangles and the second one denoted with ovals:

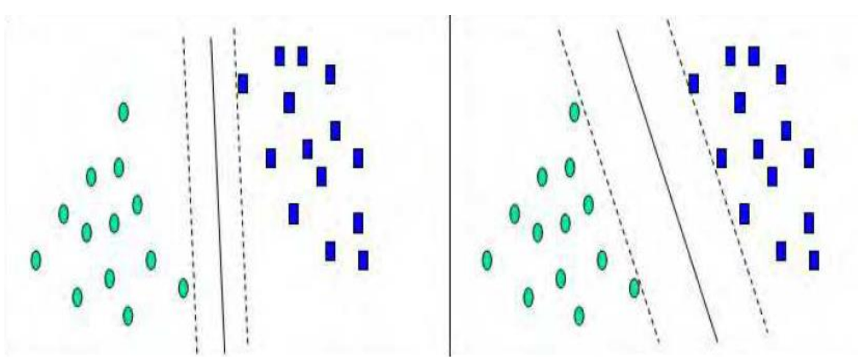


Fig:21. Two dimensions of SVM.

From above exact example we got two scenarios, first at left corner while other on the right, both cases are fully change. We separate the cases on the bases of their target categories. SVM analysis tries to get one-dimension plane for example, splits the scenarios grounded on classes. In this specific case there are immeasurable conceivable lines.
A significant thing to know is that to find the healthier line, in addition it's also important to know how we can describe finest line of all. Sunk lines strained equivalent to unravelling lines shown detachment among the in-between line and the nearby trajectories toward the line. Distance among the spotted lines is called the margin. Arguments which gratify the thickness of the boundary known as support vectors. Bellow figure demonstrates that:

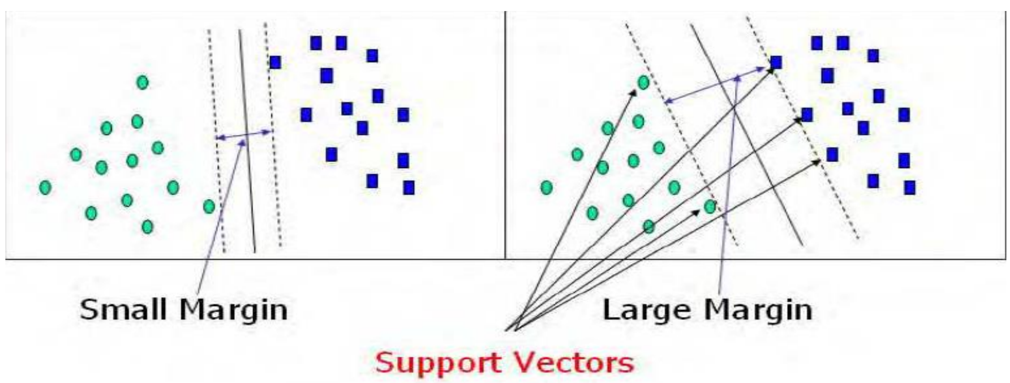


Fig:22. Support vector from Two dimensions example.

Support virtual machine investigations discover lines or hyper plane which is preoccupied, thus margin among the support vectors are browbeaten. From above picture, line of left side seems inferior from the line on right side. If every examines comprise 2 categories proposed variables having 2 interpreter variables, while group of opinions might be alienated by traditional streak. But unluckily this is not the circumstance, so SVM must work by more than 2 variables which work as predictor, secondly it will be unravelling the points with nonlinear curves, thirdly it will manage the cases in which the clusters cannot be separated fully, and fourthly it will it work on treatment classifications have categories more than two.

3.2.2 Neural Network (NN):

Job of the neural networks is to do many iterations of regression and even classification at a time, while normally each network can perform one by one of them only. in majority of cases subsequently, network would comprise variable which would belong to output only, on another hand in the specific circumstance of several state classification issue, it is maybe look like to numerous units of output

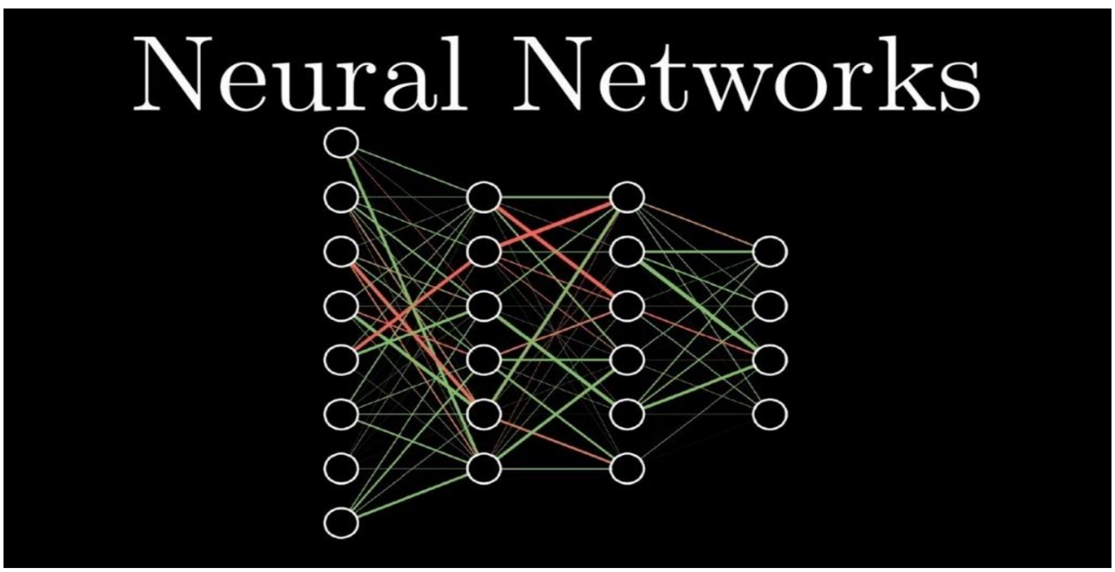


Fig:23. An Example of Neural Network:

The definition of a single network it would comprises various outputs which maybe get hurt with fraction talk. The finest answer is to skill different networks for every single output not to combined them to a cluster or group in this case they will consider as a single unit. Neural methods are: [35]

3.2.2.1 Multilayer Perceptron:

If we say t it is the most prevalent network construction in use today perhaps it will not be wrong. Using back propagation multilayer perceptron, trained as a classifier, is shown to estimate the Bayes optimum discriminant function. Outcome is recognized for both two-class problem and multiple classes. It is also made known that outputs of the multilayer perceptron estimated the subsequent likelihood functions of the classes being trained. Each unit do an incomplete sum of their own inputs then pass the level of activation on a transfer function to get its outputs. Furthermore, the feed forward topology is used to set all units layered. Network got a modest amplification of the shape of I/O model by means of the weightiness and thresholds free limits on model. This type of networks would be managing the design through the amount of layers also number of components form to each layer thoroughly with arbitrary difficulty defining the function exertion. Key complications of the multilayer perceptron design comprise arrangement of the unseen layers also the quantity of the units of layers. The problem defines by itself the number of input and output elements because of the chances that it got roughly ambiguity to making decision exactly choose the inputs need to be use. Though, for time being we are considering that the input variables are instinctively chosen and all 3them will be efficient. The hidden units to use are not so clear. The best point to start with is to use hidden layer. [36]

3.2.2.2 Training Multilayer Perceptron:

Predication error made by the network after selecting total amount of layers and the numbers of the components in every coating. Setting of the thresholds on a network is important to minimize prediction error. Old historical fashion that it adjusts automatically the weights and thresholds for minimizing the error. So, this procedure is alike to fit the model signified with the help of the training of the network by the help of available data. Weakness of using specific arrangement of network would be get by consecutively all the exercise tricks on the net one by one then compare the actual outcome made to the wanted outcomes. To give the whole network error an error function is used to gather the differences. The most common error function is the sum squared error which is specially use for the issue of regression which work in a fashion to square the error of every output and at last sum it together to get overall error. On other hand for more probabilistic cases the cross-entropy function is used. In old demonstrating methods like linear modeling the verdict of algorithmic finding the configuration of the model which would for sure reduces Error. While in greater modeling power like nonlinear models of neural networks even if we regulate network for the purpose to lesser the error in it. But it is important to note that we will not ever be so confirmed that the error is still get lower. Error surface is a useful idea in which N different weightiness and verges of the network consider as the limits of the prototypical which get as a measurement in space.

3.2.2.3 Back propagation algorithm.

The finest training algorithm in terms of neural network is back propagation. There are so many modern algorithms just like conjugate gradient which is considered much faster then back propagation, but still back propagation got so many good features that it seems efficient because it is really to understand. For now, I will overview this one, but I will also discuss progressive once later. Gradient vector which belongs to surface of the error would be calculated in back propagation. The vector point is along with a line having strident circumstantial compare to the current one, because of this reason if we move a small distance even it will drop the quantity of error. If we move in this order it will give us a minimum up to some degree. The most problematic part here is to select the size of step, its mean how long a step would be. Large steps might assemble faster but in contras it may exceeded the solution, another possibility is that if the error surface is so conservative in its nature then it may get the wrong direction. A general example for the justification of the above point from the training of the neural network is that the algorithm grows very slow with each step. On other side, it may also happen that small step size might lead us to the right direction that significant to notice that this method needs a proportion of iteration to complete its functionality. Slop and learning rate are the factor on which footstep deepened during the training. While the rate of the learning is depending on applications while its selection is based on the experiment, this might also be varying with time the growth of algorithm.

3.2.2.4 Over learning.

The core issues with the methodology defined in previous section is that, it does not reduce the error which we want to reduce it. It is the probable error which a network can possibly make with the entrance of new cases. Generally, a network is train in an away to reduces the error as much as possible based on training sets with a faultless and big data set, but it is not the similar object to reduces the mistake on the actual mistake superficial. Eventually the idea is the same.

The upmost significant appearance of this difference is the issue of over learning or may say over fitting. At ease to prove this idea of by means of polynomial arc appropriate somewhat than neural networks, nonetheless the idea is particularly the similar. If one wants to explain the data set based on a polynomial model or curve. Possibly the information is corrupt, so we didn't imagine the results to be fully precise and correct. With the use of a lower order polynomial algorithm will not work near to desired point, even if we take a higher-level polynomial it will be so flexible actually, it will arrange the data so precisely that it may take an unusual shape which will not be related to the real function. The figure given showing the functionality of such algorithm with high order polynomial sample.

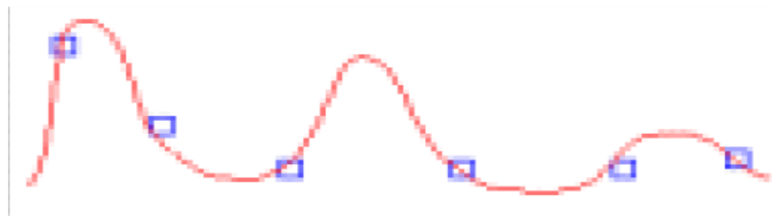


Fig:24. Sample of High order polynomial.

Same issue is with neural network even. The extra weights and too much composite function is comprising in a network, it will be considered as over fitting.
A lesser weight in a network might not be so perfectly powerful for modeling the required function.

It is an example, that fundamentally a network does design a normal linear function if it doesn't have hidden layers. The question is then how we would choose complexity of the network that is correct for our scenario? In a large Network sooner or later will habitually get a lower error; however, this type of modeling is not so good, but it may show ***over fitting***. Solution is to look at the growth in contradiction of independent available data set. There are fewer cases which are reserve, and these cases can never be use for training of back propagation approach. These cases are as a replacement for purpose to retain a free check on the development of the algorithm. In most of cases beginning network performance on selection and training sets are alike. If it's not the case and it's not same approximately the distribution of cases between the sets are partial. The training error normally get down with the growth of training, and even with providing training true error function is minimizing, in addition the selection error is also drops. On the other hand, the rise or drop in the selection error specifies that network starts to over fit the data after the end will come for the training data. The occurrence of over-fitting in this type of training process, is called over learning. It is suggested to minimalize the unseen units also the unseen layers because we made the network over powerful now.

Issue of choosing the size of the network to use and local minima, suggest that by using neural network normally contains experimenting with huge number of diverse networks, even required the training each network many times to avoid the trappings of the local minimum, and analyze every time the performance. Selection error is key to performance in this case of study, and a modest model would always be better instead of a complex model. In addition, with a very smaller improvement in error function we can select a small network instead of a larger network. The selection set play a very important role in the selection of model and this is an issue of the frequent recompences.

The self-determining guide for the performance of model lose its consistency with enough trials, because maybe you get a good network for your selection set in just few hits. For improving the sureness in the performance of concluding model, a normal practice. Final model would be testes though this dataset, the purpose is to confirm the accuracy of selection of the training sets that either its real or not.

The idea of making partitions to several subcategories is so ineffective assumed that normally we consume very minimum figures than we would rather request specially for an only subgroup. The solution to this specific issue is re-sampling. To perform trials, we use different division methods to divide the obtainable data to choose training also sets of tests.

We got a huge amount of different techniques toward the dealing with this subset, like random resampling, bootstraps, cross-validation. The successful decision design, like a best configuration neural network will be if we get it on the bases of number of tests with changed variable cases, and the remaining results would be more dependable. Whichever use those trials only to leader the choice as to which net to use, in addition with new samples we will have to train this network from base, or we should recall best network we get from the sampling process. To conclude the network plan, it consists of many stages.

- Firstly, select original network design, normally contain a hidden layer and many hidden units set.
- Perform numerous experimentations to each formation, selecting one of the finest networks founded with better selection error. Many experiments are necessary with every configuration for the purpose to get safe from being fooled.
- When under learning happens in any experiment network will never attain a fair presentation level, condition like this try to add additional neurons on hidden layers. If this idea does not help, then complement an extra layer.
- In case of over learning manifestation, the rise will come because of error function which can even remove the unseen layers.
- after getting effective configuration for your network determined from experiments with resampling to make a new network from this configuration.
- Choosing of Data: in all the overhead phases hinge on a significant guess. Particularly, the exercise, proof, and test facts obligatory to be illustrative for the original model. Earliest computer science says that garbage in, garbage out couldn’t put on extra powerfully than the neural networks displaying. The situation where if the training data is not descriptive worth of our models at the best to compromise. In worst case, it might completely inoperable. This is valuable to mention the types of issues which might corrupt the training set:

Coming isn't the history, the past is normally the training data. With the changes in the conditions, affairs from the history might not protracted. Entire possibilities would be covered. Neural network would learn firstly from the cases which exists.

Any network can be nominated and then skilled it with hundred images containing even tanks, while other which does not contain tanks. The results get by this network is 100%.

When a completely new test is conducted on new data, it proves desperate. Then what is the goal, that the pictures of the tanks are get in dim, drizzling days, and the images don't from the luminous day. A network would cram with how consider the reason for the difference among the complete bright strength. For network to work in this fashion that the network will require training scenarios which would containing altogether climate and light situations that are predictable to work with, even if all kinds of territories are not mention. Locations of the shortest distances unstable the data sets. For the meantime a network decreases a comprehensive fault, amount of diversities in the data set is dangerous. A system can be trained with the help of data set which would having 800 upright scenarios besides with 100 corrupt would prejudice their choice to decent scenarios which suits, with this fashion we make the algorithm to work and lesser the complete error which is more extremely near the decent cases.

But if exemplification of decent and depraved cases is not like in actual experiments decision from the network is might be wrong.

3.2.2.5 Self-Organized Map:

Self-organization mapping networks used different from the other networks. Though other all networks are intended for supervised learning jobs. While, the purpose of SOFM networks design are mainly to work with unsupervised learning, though in the circumstance of supervised learning data sets of the exercise hold

cases giving the input variables collected through connected outputs desirable the network essentially conclude the mapping policy between the I/O. In unsupervised learning training data set comprises just variables which belongs to input. In initial look it could look odd. Missing of the outputs, because the question arises is what would the network pick up from it?
The response is that SOFM network can efforts toward the picking up the construction of the statistics. SOFM absorb the ability to identify groups of different data, even to relate and compare different groups with each other. User would try to make sympathetic the information, that is using to recover the network. Group of the data are familiar, afterward they can be branded, and determination is that the net originate can be accomplished of organization errands.

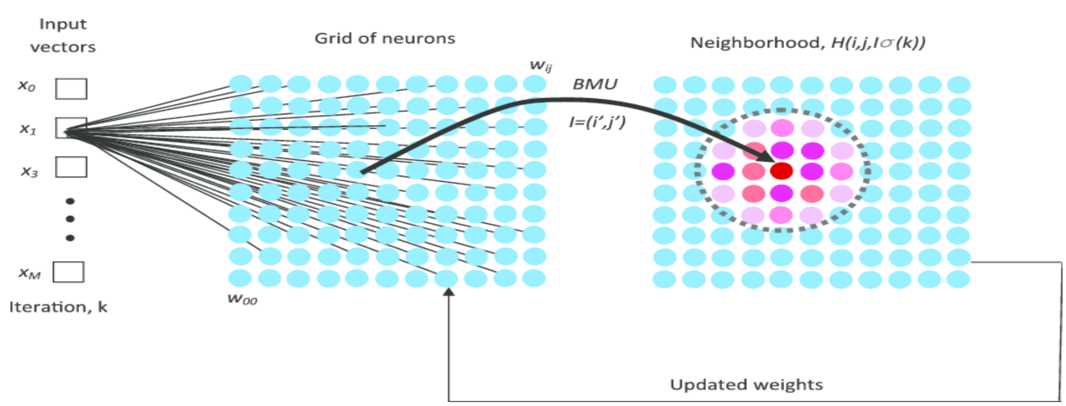


Fig:25. Self-Organization map learning:

When output groups are instantly available an SOFM type of networks would be used aimed at group in this situation, advantage of that is the aptitude to highpoint similarities among programs. Another likely usage stays in uniqueness discovery. This type of networks would to learn make out groups from data sets of the training and answer it back. Unlike previous cases if new data come across, the network will fail in recognizing it which specifies novelty. In SOFM network we got only two types of layers, the first one is input layer while the second one is output layer. [37]

Units that belongs to topological map layer is usually situated usually in two sizes in space though uniform ST Networks cares single dimensional networks. normally, the SOFM network is skilled through algorithm that is iterative.

Initially with an initial random class from circular centers. Algorithm increasingly changes the centers for the purpose to imitate the gathering the training data.

In a stage the associates the sub sampling with K means procedures use for the purpose to allocate cores in the SOM system and certainly the SOFM procedure could be used for the assignment of the centers specially for this type of networks. Though, algorithm performances with a level that is diverse. While iterative training technique is also placing in the network the purpose is that units on behalf of centers come together among the space of the inputs and even they are also located nearby to each other on map of the topology. It is a possible scenario which anyone might be think that the net's topology layers an unpolished grid with binary dimensions that can be one-sided in N different dimensional space of the inputs, the goal is to reserve the original structure as long it is possible.

Any one can try to indicate N dimensions in just two dimensions spaces, but it will cause the cost of the feature aspects. Nonetheless, the method can be valued in letting user to imagine information that maybe impossible to comprehend.

This procedure practices time function known as decaying learning the purpose is to attain weighted summation then to be sure about the modifications develop understated with the time. And this statement guarantees centers to a negotiable exemplification of different steps that cause win to the neurons.

By adding idea of the locality, the topological order of the algorithm features is attained. Locality is a cluster of different neurons neighboring by the captivating neuron. While the locality also can rotten like the rate of the learning with time. Its mean at the start several neurons would be fit to the locality that perhaps nearly the entire topological plan, though the closing ladders in the area would be nil.

Variation in the neurons is basically useful not for the captivating neuron only, but also for the associates of the present locality. The consequence of this locality apprise in the opening actual large zones of the network is dragged in the direction to the training cases also dragged very expressively.

Network grows a crude arrangement of the topology by similar cases starting groups of neurons from the topological map. With the passing of time rapidity of the learning and locality equally would be reduces therefore the healthier changes in the zones map could be haggard, finally carry about in the good alteration of the distinct neurons.

Normally, the training procedure is deliberately directed in two separate stages, one comparatively quick stage consuming strange learning speeds and locality, I in addition an extensive phase having too slight learning speed and near to zero localities. When network is skilled to categorize mapping in the information, it would be rummage-sale as an imagining means to inspect data. Datasheet of the winning incidences that counts the quantity of periods that every neuron succeeds while the cases of the training are apply and that would be reviewing to understand that if dissimilar groups have shaped on the same map. Distinct cases would be made, and topological map would be observed, the goal is to realize that if there are any cases that can be assign to clusters that naturally comprises signifying the application area that is real, as a result that the association among gathered cases can be recognized. When the groups become known the neurons of the map would be bound to display their sense with time. As soon as the topological map has been constructed, cases that are new would be forwarded to the network. Let say if the captivating neuron is considered with a specific class name the network would do sorting. But if it's not happened then the network is pragmatic as unanswered.

3.2.2.6 Data Clustering with the help of Self organize map:

In SOM the most important portion is data. In above section we mention few examples of data having 3 dimensions which are normally used in experimenting with SOMs.

Below picture shows three colors which characterized in 3 different sizes the green blue and red. Purpose of the self-organizing mapping is to scheme data having N dimensions mad about something which will be better to understand by looking to it. In this specific case we have map with two dimensions.

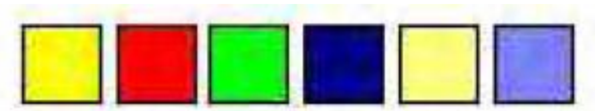


Fig:26. self-organize map colors.

We can guess that grey and dark blue will come nearby on map in above case, while red and green will come close to yellow.

The second important part of SOM is weight vectors. Every weight vector got two parts which is mentioned in the figure below. Weight vector first portion data contains by it. Dimensions is similar with the vectors that is a sample while additional part of weighted vector is usual position. Positive point of colors is that by displaying the colors we can show the data. Even in this specific case colors are the data, while x y is the location of pixels happening on screen.

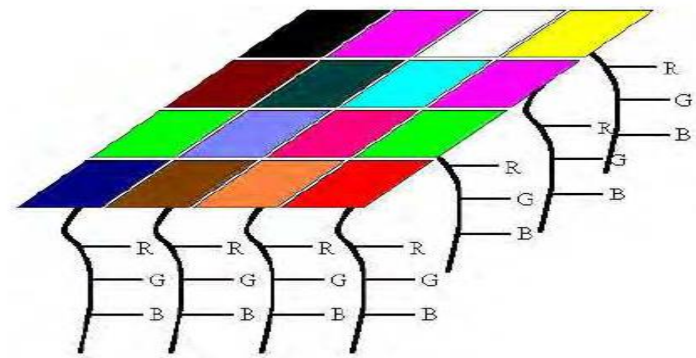


Fig:27. Weight of vector array in 2D:

The figure is alike the above figure in this sample, but weight vectors array of 2d is used. The above figure is a warped interpretation of a network that contain array of N dimensions aimed at respectively heaviness and every weight having their individual exceptional site in the network. It is not necessary for weight vectors to arrange it in two dimensions. For using SOM of one dimension a lot of work is needed. Though the part of the data from the weight vector must be get the same dimension with sample vector. Normally, weights are denoted as neurons and for now SOM in simple words are neural networks. SOMs grows by contending for depiction of the samples. By learning that tool of conversion even neurons are getting change and became as samples to compete for selection.

These learning and selection processes able the weights to get recognized to a map on basis of similarities. So, the sample and weight vectors are the two components by which, we can call the weight vectors in a proper way that we would come to know the similarities between the sample vectors. It can possible with a very simple algorithm which is the following:

Initialize Map
For t from 0 to 1

Randomly select a sample
Get best matching unit
Scale neighbors
Increase t a small amount

End for

Fig:28. An Example of the SOM Algorithm.

The primary step is to create a SOM the purpose is to start the weightiness vectors. One can choose sample vector from there arbitrarily then search chart of the vectors that contain weights to discover the best weight for representation of that sample.

There are different phases:

Initialization of the Weights:

Below there are three images showing the 3 different possibilities which decide to start map which contains weights of the vectors. The program underneath showing that there are six strengths displayed for three mentioned colors. [39]

Real values of weights are floats and its do not get away from the illustrations. So, they got a larger variety than six values which are presented in the figures below:

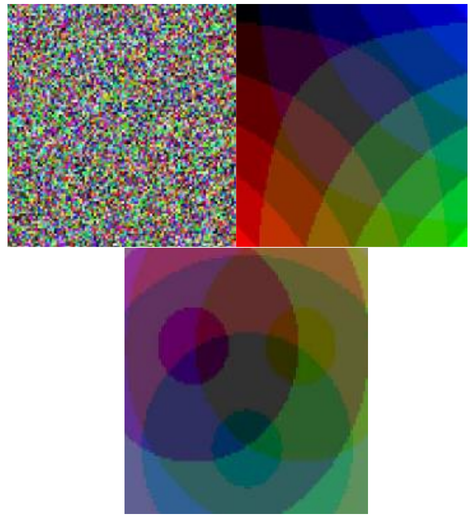


Fig:29. Weight Values.

There are different methods use to initialize weight vectors first one is to allocate a value randomly a value to each vector for the data it contains. The calculation of SOM in terms of kohon is shown in the pictures which shown on a screen that contain different pixels for different colors and in the result, we get the formation of 3 different colors, red, blue, and green. But the calculation of SOM in terms of kohon is very expensive in computation terms, but as a solution we got an alternative that is the initialization of the weight and that sample would get a specific point for its representation which would not start off too far away. In this method the number of repeating is so less then the previous and eventually it will save the time. We also having two more techniques for the initialization in adding to the accidental one. Particularly in this case just put the blue green and red and black to every corner and then fade them toward the center slightly. While in the second one all the three colors red blur and green are equally far away from center and from each other.

3.2.2.7 Getting a better matching unit.

Comparatively, in this method you need just to drive over the weight vectors in addition to calculate detachment between every weight till the sample vectors that is selected. Vectors with shortest distance would be the winner and if we get a scenario where the distance of two vectors are same then a random selection policy will be adopted between the weights having the shortest distance. For determining the distance mathematically, we got many different methods. Common is the usage of Eucidean distance.

$$\sqrt{\sum_{i=0}^{n} x_i 2}$$

X{i} i.th member value of a data sample whereas n shows the number of dimensions toward sample vectors. Proportions toward the vectors. If we take the circumstance of standards as 3D points, every module act as an axis. The value of green color is (0,6,0) while light green color is meant with (3, 6, 3) values, and these would be close to green one (6, 0, 0).

Slight green = $\sqrt{((3-0)^\wedge 2+(6-6)^\wedge 2+(3-0)^\wedge 2)} = 4.24$
Red = $\sqrt{((6-0)^\wedge 2+(0-6)^\wedge 2+(0-0)^\wedge 2)}$ = 8.49

Thus, the better matching entity is light green, nonetheless this process of computing distances and then comparing them is apply on complete map the winner in this contest will be the weight having shortest distance and the BMU.

3.2.2.8 Scale Neighbors:

Determining Neighbors:

To scale the neighboring weights, we got two parts, firstly to defining which weights are measured as neighbors, Secondly, to find how to get a weight that is much close to the vector choose as a sample. For finding the weights that are close to the winning weights we have different techniques to use. In some methods we have the use concentric squares some other use hexagon to fulfil its working principle. The one I will discuss here in my review is gaussian function, where each value will get a value grater then zero as a neighbor. I already mention above that the quantity of neighbor's decreases with time. So first, samples would transfer to a point from where it would be possibly contending for location. Present procedure almost is identical toward rough modification if it is trailed with fine-tuning. If radius of influence is getting decreases then function used for it doesn't matter at all, our concern is with decrease in radius. We just used a linear function

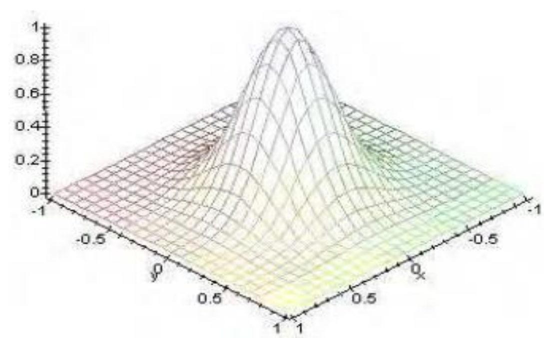


Fig:30. A graph of SOM Neighbor's determination

The plot in figure 9 showing the used function. With the passage of time the base is moving toward the center, so it's showing the number of neighbors are getting less with the passage of time. The radius at the beginning would be fix very high even few of the values would be near to the height or width of the map.

3.2.2.9 Learning Functions:

Learning function is the second possibility of neighbors discovering. Winner weight will be compensated with award to seem similar to vectors from the samples. The neighbors would also be look much similar the sample vector. One characteristics of this process is those neighbors who stay far away from the winning vector, the chances of learning will be lesser. The speed of learning weights can also be reducing, and it might be set to a value that you want. In this case we use Gaussian function by this function we get two values having range among 0 and 1 then each neighbor would be change with the help of using a specific equation. the efficient matching unit would take a T from one of its learning functions in the first iteration. And then weights will come with this process and having exactly same value from the samples that are randomly selected. The speed of leaning a weight is reduces when the neighbor's weight has fall. In first step the sample vector selected would be the winning weight till t has a range between 0 and 1. And with time to passes the winner will get slow just similar the sample where the all-out value of T reductions. The speed of learning of function will get down linearly. The nature of such function we can see in the previous plot that how the results causes the fall of bumps an its base will be shrinking. So, the neighbor will get modified with the declaration of winner weight and it will be more like sample vector.

3.2.4 Compensations and Drawbacks of SOM:

Advantages of self-organize are following:

This is an obvious thing about SOMs that it is so simple and easy to understand. If grey color connecting two of them and if they are close to each other, then we would say they are similar. While if the connecting color is black instead of grey then they are different from each other. Not like multidimensional Scaling, one can rapidly get the point that how to make then on how to use them in an efficient way. As I have mention in previous section that how they categorize data well and then easily assess the way they need it. It means one would fundamentally calculated that how decent is a map in addition it would also help to find how robust the similarities among the values are:

Following are the Drawback:

The core issues of SOMs are that how to get exact right data always. Awkwardly, to produce a map for every member of samples you need a value for each dimension which is not a good thing because occasionally it's just not possible, usually this is also so challenging to obtain the whole data, so it's preventive property about the usage of SOMs.

CHAPTER: 04

4.1 Role of ML Algorithms in the Automation of Optical Networks:

In recent past years, possibility of the precise methods in the Machine learning field have pick the consideration of academic and many researchers in optical networks and communications. The reasons of this motivation in this specific field can be classified as follow.

- **Maximize the data availability:**
  The optical networks of this modern era can provide different type of information, and for that purpose a huge number of monitors are used to provide this information from the whole system, for example, indicator of signal quality (such as bit error rate), traffic traces, different alarms e.g., specific equipment failure alarm, performance of users etc. The

contribution of ML is to predict the hidden relations between in the present different type of information.

- **Maximize complexity of the System**:
  The implementation of more advance and complicated transmission methods and introduction of more elastic principles in the networking field made the optical network more and more complex and the reason is that the increase of in the huge number of different parameters, like number of interlinks, symbol rates, modulation format, etc. The task for a network to keep all these parameters in mind and then perform is not an easy job. Furthermore, margin is also a parameter to be consider which would be adopted in the design of the model and its function is to utilize the recourses efficiently but in the result margin would increases the cost. While on another hand, Machine learning approaches conduct nonlinear models that are usually complicated, but it would mannered comparatively with relaxed training for both supervised and unsupervised method of learning that can work with information of old network data that are collected indifferent scenarios, by this simple learning approaches of ML it can solve a very complicated cross layer problems, specially I our case optical networking field. Various cases are there that get advantage from the use of ML in field of networking, in my overview I mainly divide these use cases in two categories.

4.1.1 Physical layer use case:
4.1.2 Network layer use case:

This portion of my thesis delivers an advance age overview toward the main claims of Machine Learning in networking field specially in Optical Networking. graphically presented in below picture, which make it clear why ML can be helpful in every use case.

4.1.1 In Physical Layers Terms:

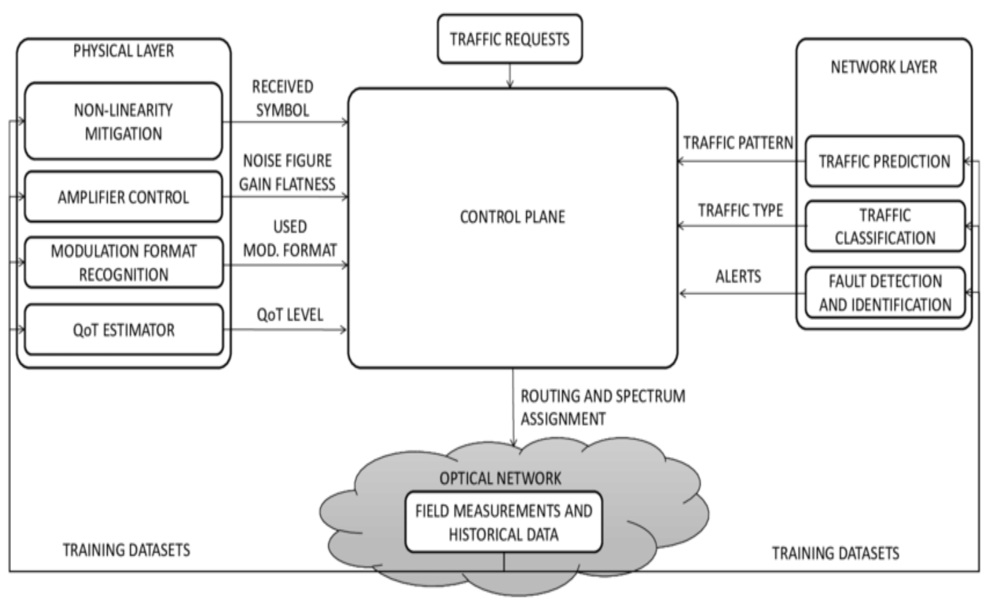
In preceding section, I mentioned that, we would deal with numerous difficulties in PL of the Optical networks. Generally, if we want to estimate the efficiency of a communication scheme then to check that if existing lightpaths are get influences by any signal degradation. Intensive care like it be used, for an example, to activate an unbeatable procedure, just like the lunch power tuning and monitoring gain of an amplifiers from an optical system. An explanation of the applications of ML is presented on physical layer is presented before any degradation in signal occurs which is difficult to recover.

- QoT estimation:

  Before the placement of the new light path, a system engineer should to evaluate the. Even for previously present lightpaths, idea of the class of communication is normally directed toward various parameters of the physical layer, for example, Bit error rate, Quality factor, and OSNR etc. and these have played an influence on supervision of the optical signal. These parameters gives a calculable ration for the purpose of knowing that if an already resolute QoT can be surefire or not, in addition it would get affected by many adjustable design constraints, like baud rate, modulation format and rate of the coding, etc. so, the optimization of this network is not so simple and when the parameters are that huge it is challenges the capabilities of designers to address manually all the groupings positioning of lightpaths.
  Meanwhile, the already existing lightpaths QoT approaches are divided in two core classes:

1) Exact analytical models which work on the estimation of physical-layer impairments, which required a long computation but the results we are getting from this class are correct.

Formulas with margin, these are faster in computation, but generally the margination rate is very high which causes underutilization of network resources. In addition, it does not have any worth because of its complex interface of various system constraints. in addition, because of nonlinear spreading of the signals in the station the derivation of correct logical model is a challenging task, and scheme face deliberation for sure can be finished to accept estimated models. Contrariwise, machine learning founds a reliable revenue to robotically guess if unestablished lightpaths would encounter the essential scheme QoT verge or not.



_

Fig:31. Overall outline of a ML-assisted optical network.

4.1.1.1 Controlling Optical Amplifiers:

Lightpath establishment is getting more and more diverse just because of the new emerging services which required very large amount of bandwidth in a very limited time. This is unfortunate, that the sever changes in the parameters of the lightpaths (like setup and teardowns) over variable wavelengths, compel the network operator to rethink on the configuration of the devices for keeping uphold the constancy of physical layer. In result of quick variations of lightpaths placement as Erbium Doped Fiber Amplifiers get grieve because of the power excursion that depend on wavelength. When a newly lightpath is added or when an already existing light path is dropped the inconsistency in the power level of different channels will depends on wavelength which are being add or drop into/from the system. Thus, an automatic controller is required before amplification signal power levels. Particularly, whenever we make cascade of different EDFAs is always crossed, to ignore this extra power of amplification divergence in different lightpaths may causes signal distortion. [40]

4.1.1.2 (MFR): Modulation Format Recognition:

ML reversion algorithms can be skilled to precisely guess power excursion before amplification. Advance optical transmitters and receivers deliver very large elasticity in different network factors just like exploited bandwidth, modulation format, and carrier frequency mostly to familiarize the bit-rate required for the transmission. In addition, it is assumed that at the side of transmission a coherent optical modulation format can be adopted. Which will capable of knowing the choice in advance which is not always the case in the receiver side, and thus it may cause affect in exact demodulation of the signals and signal dispensation and recognition.

4.1.1.3 Optical performance monitoring (OPM):

There are nine approximated analytical models which are generally use for the solution of complex non-linear problems, for capturing such effects of non-linearities supervised ML models can be design, typically for manipulating the historic statistics and making I/O bond among the checked limits and wanted output. The demand of increasing capacity requirements in the optical communication system, the most vital technique to ensure fast and consistent networks is performance monitoring. The goals of optical performance monitoring are to approximate the transmission parameters of system of optical fiber. The collected data from such parameters can be used to get rid of various goals. for example, launch power adjustment, activation of compensator modules, varying the adopting modulation format, etc. generally in optical performance it is important to collect parameters at different stages along with the lightpath.

4.1.1 In Network layer Terms:

Numerous important additional usage scenario in terms of Machine learning are emerging at the network layer. Establishment of innovative lightpaths/recovering of the present ones on the endowment of network needed very robust and composite choices which depends on numerous fast-evolving data, for example, network provider must consider the impact of the newly inserted traffic on the already existing connections. Generally, for a good and effective network operation the two most important things to be must keep in consideration are the estimation of the users and service requirement, which help to sidestep on looking after of network properties then to connected resources through satisfactory limitations on a rational price. Here is am classify the subsequent key cases.

4.1.2.1 Prediction Of traffic:

Precise guess about the network traffic in space of time lets the workers to strategy then operates its networks. Over provisioning is get decreases by prediction of the traffic as much as possible. The utilization of any sort of resource in any network operation can be improved with the help of traffic engineering which will depend on real time data.

Related Machine Learning method:

By using old collected data on user's behavior and profile of the network traffic in terms of time domain, we can skill an algorithm which will guess the forthcoming obligation of the stream of traffic and the needs of the resources. Which helps in the activation of the efficient redirecting and re-optimization of the network. which helps in putting up entirely the users flow and at same time to reduces the utilization of the resources. Even unsupervised machine learning approach can also be using to abstract mutual traffic combinations in various parts of the traffic in the network. By using this type of approach, the design and management procedures, for example, arrangement of the network capacity is stimulated likewise in numerous portions of the traffic in network. Keep in mind, that claims of traffic estimation and the comparative machine learning methods can change significantly on the bases of the well-thought-out network segment and features of traffic prediction is rely on the network segment.

4.1.2.2 (VTD) Virtual topology design:

The main idea of providing services in the transmission networks specially with the use of virtual topology is highly deployed and recommended by network experts and providers. This idea is comprising the connection of two end points same as to connect two data centers through the help of virtual link, though these two centers are not connected but the link in between is a virtual link. When all links are defined successfully and when the request for the light path is recognized, then VDT will fund a routing and wavelength Assignment issue on physical layer for each lightpath. Normally it is obvious that there must be different virtual topologies that would be present, and they would provide services to the different users. Virtual topology design use is not only to provide service to new users but also during the network failure even when the efficiency of network resources is required or when we must rearrange an already existing virtual topology. For reconfiguration of virtual topologies experts require not only the reallocating of the network capacity for a service but also needed some extra resources according to the required service or for specific feature. For example, the guaranteeing of service safety and taking the expected Qos requirement. Service provisioning like this is frequently recommended as network dividing, since every provisioned service characterizes a portion of the whole network.

4.1.2.3 Failure management:

If we would talk about the Reconfiguration of Virtual Topology Design Machine Learning classifiers would choose how to assign network resources optimally. At same time looking at many different and varied service requirements from 10 different produced computer topologies to making fast decisions and progress the provisioning of the resources specifically in situations where network can change dynamically. This ability of guessing the disappointment discovery even finding the network failure is very vital because it enables the operatives to do re-routing on time to keep facility status alive and complete Provision Level Arrangements, furthermore to recuperate from the letdown quickly. The proficiency of network management would work with different levels, like acting recognition of failure, for example, recognizing the group of light paths which are affected by a failure, is abstemiously an easy job, that allow the network workers to only reconfigure the pretentious by lightpaths. By doing so we would restore network pre-failure and that is an optimized situation in terms of utilization of resources. In addition, it also determining the reasons of the failure in a network.

4.1.2.4 Classification of Traffic Flow:

There exist dissimilar kinds of facilities in a single network then the classification of the consistent flows of traffic may allow well-organized reserve distribution, justifying the danger of over and under provisioning. Furthermore, the correct or efficient flow organization is even exploit for pre-provisioned amenities by putting on flow-specific strategies. For an example, for performing flow control and congestion control, for handling packet precedence, and for guarantying Qos to every flow based on SLAs. Bellow figure is showing the complete steps of flow classification.

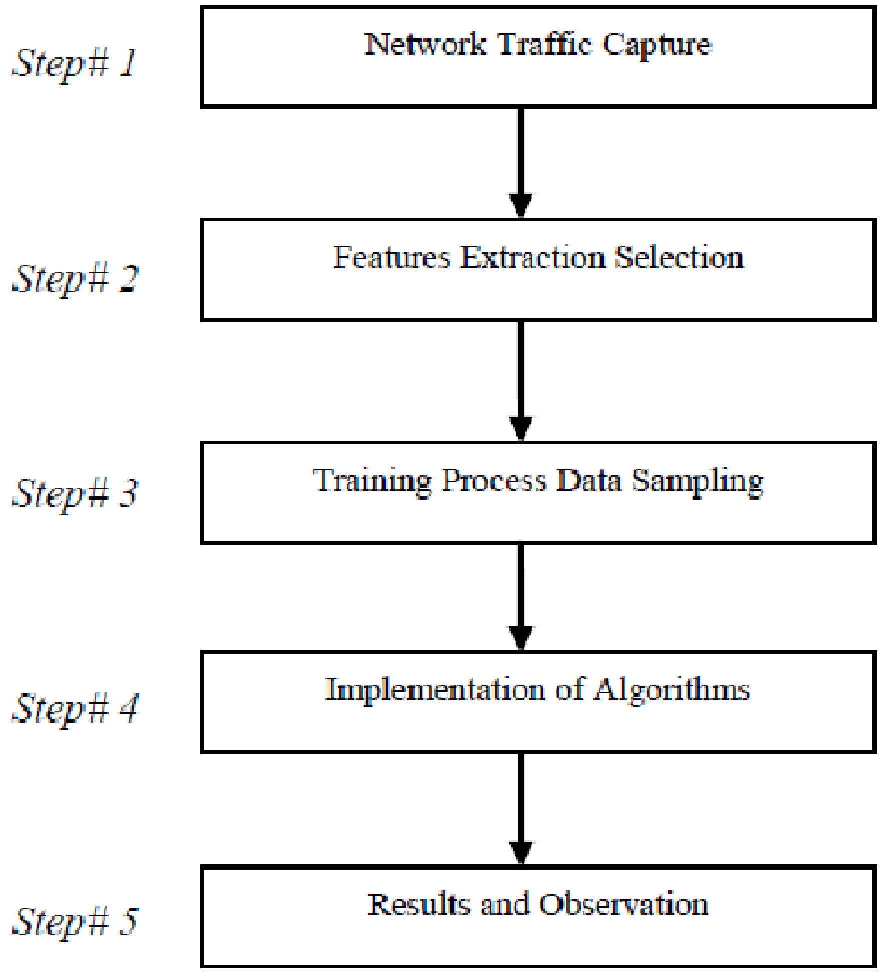


Fig:32. Network traffic classification model.

4.1.2.5 Path computation:

For a network resource provision aimed at an arriving facility appeal an appropriate path can be chosen to professionally adventure the obtainable net capitals to place the demanded traffic through the wanted Qos deprived of the distressing of the already present services that are already provisioned in the network. In general, computation of the path is did by by means of cost-based direction-finding methods.

4.1.2.5 A keen view on the surveyed studies.

Below described use cases of the physical and network layer have the capability to deal with several ML tools by the exploding of existing studies. Like, supervised/unsupervised learning, and handling dissimilar kinds of checked information of networks parameters, for example OSNR, BER, different network alarms etc. In table 1 and 2 we got the summary of the bodily and net layer use cases which point outing the topographies of machine learning methods that can be used in the literature to solve these issues. [41]

TABLE I: Different use cases at physical layer and their characteristics.

Use Case	ML category	ML methodology	Input data	Output data	Training data
QoT estimation	supervised	kriging, L_2-norm minimization	OSNR (historical data)	OSNR	synthetic
			OSNR/Q-factor	BER	synthetic
			OSNR/PMD/CD/SPM	blocking prob.	synthetic
		CBR	error vector magnitude, OSNR	Q-factor	real
			lightpath route, length, number of co-propagating lightpaths	Q-factor	synthetic
		RF	lightpath route, length, MF, traffic volume	BER	synthetic
		regression	SNR (historical data)	SNR	synthetic
		NN	lightpath route and length, number of traversed EDFAs, degree of destination, used channel wavelength	Q-factor	synthetic
		k-nearest neighbor, RF, SVM	total link length, span length, channel launch power, MF and data rate	BER	synthetic
		NN	channel loadings and launch power settings	Q-factor	real
		NN	source-destination nodes, link occupation, MF, path length, data rate	BER	real
OPM	supervised	NN	eye diagram and amplitude histogram param.	OSNR/PMD/CD	real
		NN, SVM	asynchronous amplitude histogram	MF	real
		NN	asyncrhonous constellation diagram and amplitude histogram param.	OSNR/PMD/CD	synthetic
		Kernel-based ridge regression	eye diagram and phase portraits param.	PMD/CD	real
		NN	Horizontal and Vertical polarized I/Q samples from ADC	OSNR, MF, symbol rate	real
		Gaussian Processes	monitoring data (OSNR vs λ)	Q-factor	real
Optical amplifiers control	supervised	CBR	power mask param. (NF, GF)	OSNR	real
		NNs	EDFA input/output power	EDFA operating point	real
		Ridge regression, Kernelized Bayesian regr.	WDM channel usage	post-EDFA power discrepancy	real
	unsupervised	evolutional alg.	EDFA input/output power	EDFA operating point	real
MF recognition	unsupervised	6 clustering alg.	Stokes space param.	MF	synthetic
		k-means	received symbols	MF	real
	supervised	NN	asynchronous amplitude histogram	MF	synthetic
		NN, SVM	asynchronous amplitude histogram	MF	real
		variational Bayesian techn. for GMM	Stokes space param.	MF	real
Non-linearity mitigation	supervised	Bayesian filtering, NNs, EM	received symbols	OSNR, Symbol error rate	real
		ELM	received symbols	self-phase modulation	synthetic
		k-nearest neighbors	received symbols	BER	real
		Newton-based SVM	received symbols	Q-factor	real
		binary SVM	received symbols	symbol decision boundaries	synthetic
		NN	received subcarrier symbols	Q-factor	synthetic
		GMM	post-equalized symbols	decoded symbols with impairment estimated and/or mitigated	real
		Clustering	received constellation with nonlinearities	nonlinearity mitigated constellation points	real
		NN	sampled received signal sequences	equalized signal with reduced ISI	real
	unsupervised	k-means	received constellation	density-based spatial constellation clusters and their optimal centroids	real

Fig:33. Summary of Physical and Network layer use cases.

TABLE II: Different use cases at network layer and their characteristics.

Use Case	ML category	ML methodology	Input data	Output data	Training dat
Traffic prediction and virtual topology (re)design	supervised	ARIMA	historical real-time traffic matrices	predicted traffic matrix	synthetic
		NN	historical end-to-end maximum bit-rate traffic	predicted end-to-end traffic	synthetic
		Reinforcement learning	previous solutions of a multi-objective GA for VTD	updated VT	synthetic
		Recurrent NN	historical aggregated traffic at different BBU pools	predicted BBU pool traffic	real
		NN	historical traffic in intra-DC network	predicted intra-DC traffic	real
	unsupervised	NMF, clustering	CDR, PoI matrix	similarity patterns in base station traffic	real
Failure management	supervised	Bayesian Inference	BER, received power	list of failures for all lightpaths	real
		Bayesian Inference, EM	FTTH network dataset with missing data	complete dataset	real
		Kriging	previously established lightpaths with already available failure localization and monitoring data	estimate of failure localization at link level for all lightpaths	real
		(1) LUCIDA: Regression and classification (2) BANDO: Anomaly Detection	(1) LUCIDA: historic BER and received power, notifications from BANDO (2) BANDO: maximum BER, threshold BER at set-up, monitored BER	(1) LUCIDA: failure classification (2) BANDO: anomalies in BER	real
		Regression, decision tree, SVM	BER, frequency-power pairs	localized set of failures	real
		SVM, RF, NN	BER	set of failures	real
		regression and NN	optical power levels, amplifier gain, shelf temperature, current draw, internal optical power	detected faults	real
Flow classification	supervised	HMM, EM	packet loss data	loss classification: congestion-loss or contention-loss	synthetic
		NN	source/destination IP addresses, source/destination ports, transport layer protocol, packet sizes, and a set of intra-flow timings within the first 40 packets of a flow	classified flow for DC	synthetic
Path computation	supervised	Q-Learning	traffic requests, set of candidate paths between each source-destination pair	optimum paths for each source-destination pair to minimize burst-loss probability	synthetic
	unsupervised	FCM	traffic requests, path lengths, set of modulation formats, OSNR, BER	mapping of an optimum modulation format to a lightpath	synthetic

Fig:34. Summary of Physical layer and Network layer cases.

CHAPTER: 05

5.0 DISCUSSION AND FUTURE DIRECTIONS:

This chapter is basically containing the overview about the idea of the research area of my studies. Its mean that what is the possibility of expending this area in future. And taking in concentration there are few sections which would require much higher attention in future.

5.1 Machine Learning Methodologies:

This is noticed about the current revisions that it assuming utmost of the machine learning in optical networks by using supervised learning methods that is offline. For example, the training of the machine learning algorithms is based on the preexisting survey about the issue and the collected data from existing studies and even the decisions are based on this existing study. But unfortunately, this supposition is not correct in the optical networks, wherever the situation is sharply evolving with time. Like frequent variations in the conduct of optical mechanisms and traffic variation causes issue. So, we can imagine that by learning from the samples of available past data we would implement different kinds of algorithms in machine learning just like semi-supervised learning or unsupervised machine learning would be applied to slowly get in the actual contribution data as provided by network resources. The retraining of the supervised learning is also important to investigate the mechanisms to spread the applicability to various network infrastructures. If we talk in more general sense, the newly machine learning techniques established adhoc specially for optical networking issues that are expected to be emerge in near future. For example, active machine leaning procedures, that can request operator to interact with the observed exercise data with exact physiognomies. And in this fashion the we would get a good model that will help to reduce the number of samples for making a correct model. And that for sure will reduce the cost of the network ultimately.

5.2 Availability of Data:

Now a days, service providers do not test a very big collection of data to become the realism of the results from the existing studies. This issue might be solving somehow by addressing relevant events, like signal degradation through testbeds of Optical network, but it is simply incredible to replicate the variety of the situations of actual time networks in laboratories. Even in some super situations where we have the access to the actual and accurate data of the network as mentioned in some cases above, still it is difficult to collect very large and useful datasets from a faulty operational situation, as we knew, networks are generally design and manage via traditional project methods that can made the failure likelihood insignificant.

5.3 Time scale:

Very little care has been devoted toward the point that variant applications have various timescales through which the useful data can monitored and shows in observable manner. Component behavior would become too slow with the passage of time on another hand traffic grows very fast with various time scales. The understanding of the right timescale to monitor the limits to be nourished into machine learning algorithms and it is not only significant for the optimization of the correctness of algorithm but also it is basic to design the number of controls bandwidth required to implement a scheme grounded on machine learning.

In a case where machine learning algorithms work flawlessly, nonetheless it needs an enormous quantity of data to get appraised very fast rather supplementary bandwidth control requirement would deter applications of algorithms.

5.4 Ultimate cognitive control system:

One more significant discussion about the current machine learning resolution is that it talks specifically about the remote problems in optical communications and networking. Keeping in mind that soft wear defined networking is designing to get the control upon multiple networking layers and technologies, a combined control plane like this is also known various requests of machine learning the purpose is to deliver all in one design for optical networks. Machine learning would be using to approximate various system characteristics at various layers, like QoT, occurrence of failure, pattern of traffic and some of them are dependent on each other. While, the others don't have the dependency on each other. There required some more investigation to determine the applications of based on machine learning joint frameworks for the control where overall projected parameters can be measured when creation choices for example the direction of newly routing lightpath and when would be the rerouting of an existing light path would happen and when to change the parameters of transmission for example baud rate and format of modulation. Additional auspicious and advanced area of machine learning request is failure recovery that would be correspond with SDN switch is network letdown retrieval. Claims of machine learning to this subject, with the addition to its skills to contemplate the related data across all the layers of a network, and that to move to such categorizations that are probabilistic in nature and in result that would provide much better-off data against the presently assumed limit-based strategy.

5.5 visualization:

Emerging actual tools for visualization that permit the output to be rich with information formed through ML algorithms instantly reachable and understandable for the users at the end is a main enabler for all-in-one addition of the machine learning methods in optical network organization frameworks. However, few initial investigation phases in this route has already completed. Whereas the bubble charts and spectrum color maps are deployed to look at the links of the network facing higher bit error rate, strategy rules for instinctive conception methods conditional on exact goal of ML practice. [42]

5.6 Standardization and Commercialization:

Still the applications of ML methods are in upmarket to optical networking which have now tangled the attention of the operators of networks and the vendors of the optical apparatus. Furthermore, it is projected that this courtesy would raise very fast. Some actions are observed basis on QoT to guess the level of optimization to decreases the margin and making it attentive if the error already on low margin optical network policy among others on guessing of traffic and irregularity discovery. In addition, standards that are accountable for standardization they have determined observing feature for determination of networking difficulties. However, it is the most we know that there is no explicit action is undertaking in current time that dedicated courteousness on optical networks. And finally, a speculative area of investigation in future in optical networks is optics for machine learning, due to its nonlinear nature the optical components will connect in way to make learning situation. This method shows all optical alternate for traditional software operations. For the creation of photonic neural network, the we use semiconductor laser diodes through time multiplexing getting benefit of non-linear reaction toward power injection because of the joining of amplitude and segment of optical field. Reservoir computing is a method of machine learning which would applied through a nonphotonic reservoir that consist of a network of joined crystal cavities. And this behavior of resonating is very helpful because the power would be stowed in the hollows and produce belongings that would be nonlinear, and the network would train to regenerate episodic patterns. In nutshell, applications of machine learning to optical networking is a very rapidly rising research subject which is facing an enormous contribution from industry as well the academics. [43] In my thesis, I have just tried to give a petite conversation on the conceivable future directions and its for sure that in future there must be so many topics that would emerge.

CHAPTER: 06

Conclusion:

In this advance time Machine learning is play a very vital role in almost every industry, and there are a lot of research going on around on industrial level as well as on academic level about machine learning. Here I would try to give some already in used applications of Machine learning in different areas in industry. For example, Netflix Recommendation system, AlphaGo, Drug Discovery, Character recognition, voice assistance, assisted driving, or self-driving cars, face detection(Facebook and many others are using this to defect and recognize faces), Cancer diagnosis(IMB Watson is using this, which is based on machine learning), Optical networks(Almost all vendors are working to deploy ML in Optical networks) in short machine learning is everywhere nowadays.

To explain a learning algorithm in simple words. let's say, if I give you a set of data points and then I ask you to tell that how the computer will essentially set a curve to that data point and in some then somehow learn a model for that data, and then it could use to predict behavior in other situations. The above example is quite closer to what I want to explain about the machine learning algorithm. We would like to have a program that can learn from its experiences, something that can deduce the new facts. I would like to present a simple but attractive definition of ML by "Arthur Samuel" in 1959, according to him "ML is a field of study that can gives computers that ability to learn without being explicitly

programmed" and I think he wrote the first program because he also learns from experience.
My main goal here is to explain how a system can learn without been explicitly programed. One way to think about it is to think about how we would normally programmed and how and what we would like from a ML algorithm, I knew nobody would agree upon normally programming because a program would never be normal but let say traditional programming, so in traditional programming the process is that I write a program that I can put to the computer and then it can take data and produce some appropriate output. But in machine learning scenario things are different we would have to give the computer an output from previous experiences, that what I want, and the task of the computer is now to produce a program based on provided previous results as an input to this system. So, this act as a loop, machine learning is using the output of previous system as an input and in results provide us a useful new program. It means we don't have to program everything, but the system will manage it by its own by the assistance of ML. So, one can say ML is the imperative way of knowledge that can deduce new facts from the old once, in other words it can assume that the past predicts the future. For the experiments we would give some training data to the system and based on that training data we would skilled the system and it would be capable of predicting the future. So, the basic structure would be in three main steps:
1: Training data: Observe set if examples.
2: Infer something that about process that generated that data.
3: Test data: Use the inferences to make predictions about previously unseen data.

Optical networks are been spreading and growing very smartly after the introduction of the SDN intelligible transmission. An SDN platform is used to combined progress in the direction of high-performance hardware and smart software which and this SDN platform provides a compact base for guaranteed novelties in optical networking.

Progressive machine learning algorithms can use the large amount of data obtainable from network elements and to make them learn by experiencing and to style the network in more extra adaptive fashion. The exploration of the applications of machine learning it already pick the concentration of the researchers for enabling a clever optical network. And in my thesis, I have tried to summarize few of the possible machine learning algorithms that is useful in making an optical network automated and I gave a try to provide keenly some possible research directions in near future. As I already mentioned that we are dealing two kind of learning, labeled and Unlabeled, now I just want to clarify both with an example, but before that I want to mention some important issues of concern with the learning of model, a lean model will depend on.

1: Distance matric between examples.

2: choice of feature vector.

3: constraint on complexity of model.

- Specified number of clusters.
- Complexity of separating surface. Want to avoid over fitting problem. (each example is its own cluster, or its own separating surface)

Machine learning is so hot in communication because of the classification and regression properties machine learning make it possible the automation of the networks. The example of classification in terms of telecommunication is symbol detection. Like QPSK and decision boundaries that classify them into four groups. In addition, we also doing very frequently the Joint carrier frequency and phase offset estimation. The foundation of machine learning algorithms in optical communication are sensors, abstraction, algorithms, and actuators. And the working principle is also same as above mention examples of machine learning. The main approach is to construct nominal model (ask for the baseline), Assume that data are far from the baseline are generated by different process(anomalies). Evaluation of far is usually require a hyper-parameter. And we got a lot of ways to do this. In networking world this automated network is also known by another name called self-driving networks. It means that network would be skilled from the previous data flows and programs going on it and it would skill itself for the future decisions that would take it by its own. And automation and autonomy are based on AI that enable the network designers to tackle that never ending trade-off we have between performance and agility. Where performance refers to traditional speed and feed. while Agility, is the speed at which you can get infrastructure to do something that you did not anticipate. And this the actual goal of AI in networking area to make networks autonomous, which mean to discover things by network itself, also network will configure by itself, furthermore network will monitor by itself, self-correction, detection automatically any problem, and auto establishment customers, Etc. The budget for making a network completely with above features is unknown yet. That would be to allow people to deal with the higher level like new service strategy. Even preventive service formation. Debauched, smart reply to safety parameters. The test for the network engineers of this age is to track a data center for 6 months or more without any human interference with no reduction and compromise in the performance and functionality.

References:

[1] Epistemology and Knowledge Management Concepts And Practices. Journal of Knowledge Management Practice. Allix, N. M. (2003, April).
[2] Introduction to Machine Learning. Massachusetts, USA: MIT Press. Alpaydin, E. (2004).
[3] Learning and Memory. Wiley, New York, USA. Anderson, J. R. (1995).
[4] Socialization influences on preparation for later life. Journal of Marketing Practice: Applied Marketing Science, 5 (6, 7, 8), 163 - 176. Anil Mathur, G. P. (1999).
[5] Design of a Brain, the Origin of Adaptive Behavior. John Wiley and Son. Ashby, W. R. (1960).
[6] an Analysis of Four Missing Data Treatment Methods for Supervised Learning. Applied Artificial Intelligence, 17, 519-533. Batista, G. &. (2003).
[7] Neural Networks for Pattern Recognition. Oxford, England: Oxford University Press. Bishop, C. M. (1995).
[8] Pattern Recognition and Machine Learning (Information Science and Statistics). New York, New York: Springer Science and Business Media. Bishop, C. M. (2006).
[9] The Perceptron: A Model of Brian Functioning. 34 (1), 123-135. Block H, D. (1961).
[10] Introducing Neural Networks. Wilmslow, UK: Sigma Press. Carling, A. (1992).
[11] Machine Learning, Neural and Statistical Classification. Prentice Hall Inc. D. Michie, D. J. (1994).
[12] Fundamentals of Neural Networks. New York: Prentice Hall. Fausett, L. (19994).
[13] The strange story of the Perceptron. Artificial Intelligence Review, 4 (2), 147-155. Forsyth, R. S. (1990).
[14] A learning machine: Part, 1. IBM Journal, 2-13. Friedberg, R. M. (1958).
[15] Unsupervised learning algorithms are designed to extract structure from data. 178, pp. 1-8. IOS Press. Ghahramani, Z. (2008).
[16] Artificial Intelligence and Scientific Method. OUP Oxford.
[17] Neural Networks: A Comprehensive Foundation. New York: Macmillan Publishing. Haykin, S. (19994). Gillies, D. (1996).
[18] A Survey of Outlier Detection Methodologies. Artificial Intelligence Review, 22 (2), 85-126. Hodge, V. A. (2004).
[19] Adaptive Algorithms for Discovering and Using General Patterns in Growing Knowledge Bases Policy Analysis and Information Systems. Holland, J. (1980).

[20] Experiment in Induction. Hunt, E. B. (1966).
[21] Data Mining Practical Machine Learning and Techniques (Second edition Ed.). Morgan Kaufmann. Ian H. Witten, E. F. (2005).
[22] Jaime G. Carbonell, R. S. (1983). Machine Learning: A Historical and Methodological analysis. Association for the Advancement of Artificial Intelligence, 4 (3), 1-10.
[23] Self-Organizing Maps. Kohonen, T. (1997).
[24] Unified dual for bi-class SVM approaches. Pattern Recognition , 38 (10), 1772-1774. Luis Gonz, l. A. (2005).
[25] A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophysics, 115-133. McCulloch, W. S. (1943).
[26] The Discipline of Machine Learning. Machine Learning Department Technical report CMU-ML-06-108, Carnegie Mellon University. Mitchell, T. M. (2006).
[27] Learning Language in Logic. In L. N. Science, Learning for Semantic Interpretation: Scaling Up without Dumbing Down (pp. 219-234). Springer Berlin/Heidelberg. Mooney, R. J. (2000).
[28] Transforming declarative advice into effective procedures: a heuristic search cxamplc In I?. S. Michalski,. Tioga Press. Mostow, D. (1983).
[29] Principles of Artificial Intelligence (Symbolic Computation / Artificial Intelligence). Springer. Nilsson, N. J. (1982).
[30] Evolving Evolutionary Algorithms Using Linear Genetic Programming 13 (3), 387 - 410 . Oltean, M. (2005).
[31] Orlitsky, A., Santhanam, N., Viswanathan, K., & Zhang, J. (2005). Convergence of profile-based estimators. Proceedings of International Symposium on Information Theory.
[32] Proceedings. International Symposium on, pp. 1843 - 1847. Adelaide, Australia: IEEE.
[33] Artificial Neural Networks. Singapore: Prentice Hall. Patterson, D. (19996).
[34] Learning from Observation: Conceptual Clustering. TIOGA Publishing Co. R. S. Michalski, T. J. (1983).
[35] Probabilistic Models of the Brain. MIT Press. Rajesh P. N. Rao, B. A. (2002).
[36] Mathematical Biophysics: Physico-Mathematical Foundations of Biology Chicago: Univ. of Chicago Press. Rashevsky, N. (1948).
[37]) Pattern Classification. Richard O. Duda, P. E. (2000
[38] Reinforcement Learning. MIT Press. Richard S. Sutton, A. G. (1998).
[39] Pattern Recognition and Neural Networks. Cambridge University Press. Ripley, B. (1996).
[40] The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review, 65 (6), 386-408. Rosenblatt, F. (1958).

[41] Artificial Intelligence: A Modern Approach (2nd Edition ed.). Upper Saddle River, NJ, NJ, USA: Prentice Hall. Russell, S. J. (2003).
[42] Machine Learning: An Artificial Intelligence Approach (Volume I). Morgan Kaufmann. Ryszard S. Michalski, J. G. (1955).
[43] Machine Learning: An Artificial Intelligence Approach. Ryszard S. Michalski, J. G. (1955).