



POLITECNICO DI TORINO

DEPARTMENT OF CONTROL AND COMPUTER ENGINEERING

Master degree course in Computer Engineering  
(Data Science)

Master Degree Thesis

**Penalization model in  
quantitative analysis of human  
fine motors functions**

**Supervisors**

Prof. Paolo Garza

PhD. Sven Nõmm

**Candidate**

Luca TRIGILI

matricola: 252577

TURIN October 2019

I hereby declare that, the contents and organisation of this thesis constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

.....

Luca Trigili  
Turin, 14 October 2019

This work is subject to the Creative Commons Licence

## Sommario

I movimenti fini sono ampiamente usati in psicologia e neurologia per diagnosticare i disordini cerebrali e le malattie neurodegenerative come il Parkinson e l'Alzheimer [18]. Gli ultimi dieci anni di ricerca hanno prodotto una piattaforma solida di test digitalizzati, che permettono di registrare e analizzare molti parametri di massa che non potrebbero essere osservati a occhio nudo. Tuttavia l'applicabilità dei test sui movimenti fini per modellare l'avanzare della malattia rimane una questione aperta. Una delle principali preoccupazioni riguarda l'affidabilità dei risultati prodotti dai test, che per i movimenti grossolani è stata relativamente ben studiata. Al contrario, i movimenti fini ed in particolare gli esercizi di disegno e scrittura non hanno ricevuto il medesimo livello di attenzione. I risultati prodotti dalla presente tesi descrivono le relazioni tra differenti feature che si riscontrano nei test dei movimenti fini di soggetti sani osservati per differenti archi temporali. I modelli ricavati costituiscono le basi necessarie al fine di conferire maggiore affidabilità ai risultati dei test.



## **Abstract**

Fine motor tests are widely used in psychology and neurology to diagnose brain disorders and neurodegenerative diseases like Parkinson's and Alzheimer [18]. The last ten years of research have produced a solid platform of digitized tests, allowing to record and analyze many motion parameters which could not be observed by the naked eye. Nevertheless, the applicability of the fine motor tests to model progress of the disease remains an open question. One of the main concerns is the reliability of the the test results. For the gross motor motions, the reliability of results is relatively well studied. At the same time reliability of results for the fine motor motions, and especially for the drawing and writing tests did not get the same attention. Research results reported in the present thesis describe the relationship between different features occurring in fine motor motions of healthy test subjects observed during different periods of time. Constructed models constitute the basis required to have more reliable results.



# Acknowledgements

And I would like to acknowledge Sven Nõmm from Tallinn University of Technology, who supported the development of the research performed in order to reach the conclusion of this thesis.

# Contents

<b>List of Tables</b>	<b>6</b>
<b>List of Figures</b>	<b>7</b>
<b>1 Introduction</b>	<b>9</b>
<b>2 Literature overview</b>	<b>13</b>
<b>3 Background</b>	<b>17</b>
3.1 Luria's tests . . . . .	17
3.1.1 Test types . . . . .	18
<b>4 Experimental settings</b>	<b>21</b>
4.1 Hardware . . . . .	21
4.2 Software . . . . .	21
4.2.1 Data description . . . . .	21
4.2.2 Data acquisition . . . . .	23
4.3 Research and development . . . . .	23
4.4 Mathematical method . . . . .	24
4.4.1 Statistical one population sample test . . . . .	24
4.4.2 Linear regression . . . . .	25
4.4.3 Cross-validation . . . . .	26
<b>5 Features</b>	<b>29</b>
5.1 Feature extraction . . . . .	29
5.2 Feature categories . . . . .	30
5.2.1 Pressure, geometric and kinematic features . . . . .	31
5.2.2 Drawing features . . . . .	31
5.3 Feature naming . . . . .	32



5.3.1	Feature details . . . . .	33
5.4	Feature List . . . . .	35
<b>6</b>	<b>Formal problem statement and Methodology</b>	<b>37</b>
6.1	Statistical analysis . . . . .	38
6.1.1	Statistical analysis on test exercise performed in different sessions . . . . .	38
6.1.2	Statistical analysis on individual test exercise . . . . .	41
6.2	Model building . . . . .	43
6.2.1	Data preparation . . . . .	45
6.2.2	Backward feature elimination . . . . .	45
6.2.3	Model validation . . . . .	47
6.2.4	Modelling individual test exercise . . . . .	48
<b>7</b>	<b>Result</b>	<b>49</b>
7.1	Statistical analysis . . . . .	49
7.1.1	Statistical analysis on test exercise performed in different sessions . . . . .	49
7.1.2	Statistical analysis on individual test exercise . . . . .	52
7.2	Model building . . . . .	56
7.2.1	Modelling test exercise performed in different sessions . . . . .	56
7.2.2	Modelling individual test exercise . . . . .	58
<b>8</b>	<b>Conclusion</b>	<b>63</b>
	<b>Bibliography</b>	<b>65</b>

# List of Tables

4.1	Test values table . . . . .	23
5.1	Feature naming table . . . . .	32
7.1	Features — STATISTICAL ANALYSIS — PLtrace — Weekly — first and second session . . . . .	50
7.2	Features — STATISTICAL ANALYSIS — PLtrace — Weekly — first and last session . . . . .	50
7.3	Features — STATISTICAL ANALYSIS — PLcontinue — Weekly — first and second session . . . . .	50
7.4	Features — STATISTICAL ANALYSIS — PLcontinue — Weekly — first and last session . . . . .	51
7.5	Features — STATISTICAL ANALYSIS — PLtrace — Daily — first and second session . . . . .	51
7.6	Features — STATISTICAL ANALYSIS — PLtrace — Daily — first and last session . . . . .	52
7.7	OLS Regression Results with acceleration_nc as dependent vari- able . . . . .	59
7.8	OLS Regression Results with azimuth_nc as dependent variable	61

# List of Figures

3.1	Pattern PL . . . . .	18
3.2	PL Continue example . . . . .	19
3.3	PL Trace example . . . . .	19
4.1	Least squares error approach . . . . .	26
4.2	Diagram of k-fold cross-validation with k=4. . . . .	27
5.1	Area Difference . . . . .	32
6.1	Method to compute differences between features . . . . .	39
6.2	PL series divided in intervals . . . . .	41
6.3	Linear model building process . . . . .	43
6.4	Matrix representation of the observation . . . . .	45
6.5	Backward feature elimination process . . . . .	46
7.1	PLtrace - Weekly - Differences between sessions 1-2 Cumulative True . . . . .	53
7.2	PLcontinue - Weekly - Differences between sessions 7-8 Cumulative True . . . . .	54
7.3	PLtrace - Daily - Differences between sessions 4-5 Cumulative False . . . . .	55
7.4	PLcontinue - Daily - Differences between sessions 4-5 Cumulative False . . . . .	56
7.5	$R^2$ adj. over number of intervals . . . . .	57
7.6	Common patterns between group of subjects . . . . .	58
7.7	Prediction of acceleration_nc based on pressure features . . . . .	60
7.8	Prediction of azimuth_nc based on pressure features . . . . .	61



# Chapter 1

## Introduction

The present thesis providing the basis required to apply corrections to the test result in case of anomalies or cheating by the Parkinson's disease (PD) patients. A growing part of human population suffers from a degenerative disorder of the central nervous system that mainly affects the motor system (e.g. PD). Recent studies confirm that in Italy PD affects approximately 500 people per 100.000, which amounts to more than 320.000 people [1]. These evidences put Parkinson's disease on the list of the most common neurodegenerative disorders. Parkinson's disease is a progressive nervous system disorder that affects movement. PD severely affects the quality of life of the patient. There is no known cure at the moment, but early detection of the disease could drastically change the development of the illness. Symptoms start gradually, usually starting with a barely visible tremor in one hand. Tremors are common, but the disorder also induces rigidity or slowing of movement. The face may show little or no expression in the early stage of PD. The speech may become soft or slurred. Symptoms worsen with time passing. Even though Parkinson's disease cannot be cured, early diagnosis and proper treatment may significantly improve the condition [2]. Recent research focuses on gross and fine motor functions. The present thesis has its attention onto fine motor analysis. To study fine motor function, the researcher chose handwriting exercise, because is a complex task which involves a lot of muscles, and high precision movement. This process is gravely disrupted by PD and for this reason, could be taken as a good biomarker to diagnose the illness. There are also some results that demonstrated that fatigue, especially mental fatigue, can affect the performance during the tests [6]. Respective literature shows

that “Luria’s alternating series fine motor tests” permit diagnosis of PD. According to Luria, an intentional movement is the outcome of the planning process with many levels [3]. It expects a specific goal to be accomplished. The first stage is the general planning; this level answers the question of why and how some action should be performed. On the second level, concrete motion patterns are generated based on the general plan. These motion patterns are referred to as motion melodies [6]. Motion melodies are the sequences of motions, ordered in time, which should allow the accomplishment of the goal. On the third level “orders” are generated in the direction of the spinal cord. On this level, the melody of the motions is implemented. Digital Luria’s alternating series test is selected for this purpose. Luria’s alternating series tests have been used by the medical community for a long time, especially by neurologists and psychologists. Until the present, all those tests were mainly conducted with paper and pen. The patient was evaluated by the doctor with no guarantee on the objectivity of the result not perfect due to human limits. Digitization of fine motor tests is still in its infancy and needs to gain acceptance among the medical community. The slow progress may partially be caused by the conservative nature of the medicine. Nowadays a lot of studies have been done on the digitalization of different handwriting tests with the use of touch-screen technology such as drawing star [7], spiral [7], circle [7], clock [8], and various geometric figures [9]. Other studies have been done on sentences and character sequences [10].

It is a well-known fact that results of the consequent fine motor testing may be affected by the adaptation to the test [4]. While this fact is accepted by the medical community there are very few results available on this subject. And there is no properly described penalization model. This research aims to investigate the influence of adaption on the fine motor testing results in two groups of healthy subjects. Subjects in the first group were tested once a week and subjects in the second group were tested on a daily basis. It was demonstrated that weekly testing produces a minor effect on the testing results whereas it was not possible to construct a penalization model. For those tested on a daily basis, the results clearly pointed out the following: it is possible to construct penalization models. Among the subjects, it is possible to distinguish different subgroups of individuals with respect to the adaptation patterns. The last finding is in line with the results available in the literature [4]. The penalization model will constitute the basis required to apply corrections in motor functions caused by anomalies or cheating that can assist

the personnel in the diagnosis of the patient. Also, software developed in the present research is fully functional and may be used to continue research on this problem.

The thesis is divided into nine chapters as follows. *Chapter 2* introduces all the relative research. *Chapter 3* consists of the background needed to understand the presented work. *Chapter 4* provides a high-level overview of hardware and software. This chapter also briefly describes the mathematical methods used in the thesis. *Chapter 5* defines features. *Chapter 6* formalizes the problem treated in the present research. *Chapter 7* explains the methodology used. *Chapter 8* present the results. *Chapters 9* discuss the results and the possible evolutions of the presented work in the future.





## Chapter 2

# Literature overview

According to the vast majority of research papers, “micrographia” is the most common Parkinson’s disease-specific handwriting and drawing symptom [11, 12]. Which consists of a reduction in writing amplitude. Some authors [13, 14] divide micrographia into two types. “Consistent” when a uniform reduction in letter size is observed, compared to writing before PD was diagnosed. The other type is “progressive” when the subject is not able to maintain the initial size of characters writing consecutive letters. It is not difficult to detect Micrographia because researchers deal with the noticeable result of handwriting. Drawn objects and characters size are easily measured, but are those measures significant features of PD handwriting? With touch screens and tablets nowadays technology allows researchers to precisely measure pen coordinates simultaneously with time, allowing distinguishing of new kinematic features of handwriting such as acceleration, duration, velocity, jerk, slope. Some studies propose new term “dysgraphia” [13] (from prefix “dys” in medical language “disordered”), to easily explain motor aspects of the condition: akinesia (absence of power in movements), trembling, rigidity and slowness and their connections with kinematic features.

Recent studies [16, 15] used Apple Pencil stylus with an iPad touch screen and prove it more than capable of capturing precisely handwriting data. Also, iPad technology introduces two new measures, altitude and azimuth angles [16] of the Apple Pencil that may provide more useful features describing PD handwriting. Even a common smartphone touchscreen has been utilized by Aghanavesi et al. [17] for spiral drawing tests and successfully obtain kinematic features, which contain important symptom information for detecting and assessing PD coordination. Furthermore, traditional “pen and paper” tests

should be covered. The vast majority of the studies done in the past were carried out with such a method well. In recent studies, Raudmann et al. [18] uses pencil and paper tests for writing simple and more complex sentences to understand how PD handwriting changes from healthy controls (HC<sub>s</sub>). Research confirms that the writing of PD patients differs clearly from HC<sub>s</sub>. Principally micrographia and slowness in the movements were observed. A diversity of drawing tasks can be noted in the test methodology. Smits et al. [7] presented a collection of standardized tasks, which include drawing of spiral, star, circle and writing of sentence and letters sequence “elelel”. Drotàr et al. [19] conducted several studies [21, 20, 22] adopting a set of handwriting tasks, which consists of single characters, bi-grams, tri-grams, single words, sentence, and Archimedean spiral. Letanneux et al. [13] also suggests taking into consideration some aspects while choosing methodology. The scope of the drawing tests is to focus on a real low level of motor functions so tests where the subject has to write complex sentences and uncommon terms or also drawing elaborate shapes should be avoided because it implies the cognitive process. Nõmm et al. [9] proposes digitalized version of “Poppelreuter’s overlapping figures test”, which is used in psychology and neurology to assess visual perceptual cortex function [23]. Nackaerts et al. [24] created the “Systematic Screening of Handwriting Difficulties” test (SOS-test), where subjects were examined two different times within month period. The patients were invited to copy as much as possible of a text within 5 minutes with the instruction to write as neatly and quickly as in daily life. Korner et al. [25], Souillard-Mandar et al. [25], Brodaty, and Moore [8] examined “Clock Drawing Test” or CDT, where participants are required to draw the face of a clock, mark in the hours and then draw the hands to show a defined time. This method had proved its medical efficacy as a screening tool for cognitive disorders, such as Alzheimer’s and Parkinson’s disease or dementia. In Denmark, CDT is recommended [25] as a screening method to evaluate renew request of driving license made by aged people, after their 70th birthday. Abstract shapes and geometric [9], including spiral [26] and Luria’s patterns [3], are an immeasurable source of kinematic features, which help to assess tremor, bradykinesia, and overall dysgraphia. With sentence and letter sequence writing, features related to micrographia can surely be created [7, 24]. Pinto and Velay [14], Letanneux et al. [13] validate, that the preponderance of present research identifies the most significant features for PD handwriting as size, duration, velocity and writing fluency (variations in velocity, acceleration, and jerk). The features that are

not associated with micrographia are frequently named kinematic features. Nonetheless, other feature types can be found in different studies. Drotàr et al. [21] suggest pressure amounts of the Wacom tablet as good features to be utilized, which were investigated along with kinematic features, such as velocity, acceleration, and duration and manifested notable discrimination power. Drotàr et al. [20] also successfully used in-air measurements of the Wacom tablet pen that is able of recording up to 10mm height, analyzing the subject's movements linking singular strokes. Moreover, modern research [16] gets the advantage of new azimuth and latitude angles of iPad pencil for the CDT test. "Composite Index of Speed and Pen Pressure" or CISP was introduced by Zham et al. [27] and investigate its correlation with UPDRS (The Unified Parkinson's Disease rating scale) and pointed out, that CISP of spiral drawing indeed strongly correlates with UPDRS. The present thesis is related to a bigger research series at Tallinn University of Technology, which examines human handwriting and drawing. Previous works involve: "Quantitative analysis of the kinematic features for the Luria's alternating series test" [3, 28] and "Digital Clock Drawing Test Implementation and Analysis" [16].



## Chapter 3

# Background

Recent studies determine that PD affects the amount and smoothness of the human motions [30]. The changes in amount and smoothness of the motion are reflected by the values of features introduced in section 5.2. Luria's alternating series tests is a known technique [3, 29] designed especially to expose changes caused by neurological disorders on every phase of motion, execution and planning during handwriting exercise.

### 3.1 Luria's tests

According to Luria's research [29], human motor function, is the result of a complex multilevel procedure. Each complex motion requires several phases:

1. **Motion Planning Phase**: on this level, decision to take certain action is made.
2. **Motion Pattern Generation**: on the second level detailed motion pattern is generated. Motion pattern can also be described as a series of actions ordered in time. These motion patterns are referred to as motion melodies [3].
3. **Motion Execution Phase**: motion melody is being implemented when melodies are translated in signals to the spinal cord.

### 3.1.1 Test types

Luria studies lead a battery of different alternating series tests, which involve the repetition of a pattern. Luria's alternating series have different drawing patterns. In this research, the 'PL' pattern was selected. The explanation behind such naming is in the pattern which resembles the Greek letters Pi ( $\Pi$ ) and Lambda ( $\Lambda$ ).

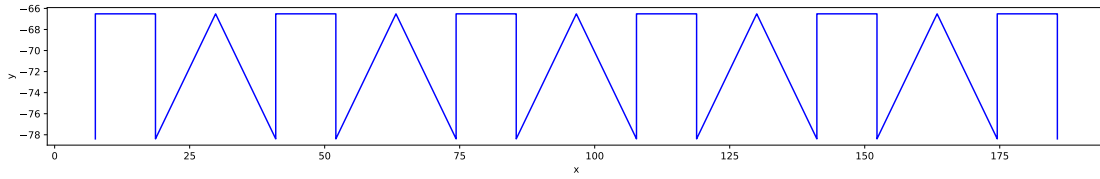


Figure 3.1. Pattern PL

For the scope of the present thesis the following tests were chosen:

- **Lines Test**: require tested subject to draw three straight lines on the screen. Test is needed purely to have a reference on the amount of pressure used by the tested individual. The test was kept to preserve consistency with other studies and give the possibility for others to use the datasets collected.
- **Spiral Test**: there are multiples way to conduct this test. The one chosen require tested subject to draw a spiral from outside to the inner part, staying between the border of a displayed pattern reference image. This test was kept because it is very popular in this field of research [7] and is helpful for comparison with the other results available in the literature.
- **$\Pi\Lambda$  Alternating Series Tests**:
  1. **Continue Exercise**: requires the tested subject to continue drawing a pattern from a few visible segments. This is the most difficult task both motion execution and complex planning processes are required. This may cause difficulties for healthy subjects, since reference drawing is not fully shown, possible borders of the required pattern are not obvious. This exercise requires the subject to switch between two different activities. Difficulties in the shifting of cognitive

sets and perseverative behaviour have been shown to be part of the neuropsychology of Parkinson's disease.

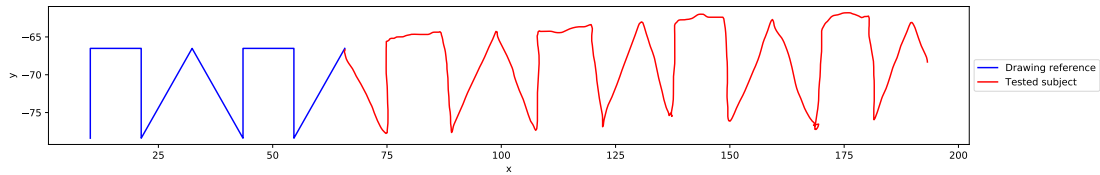


Figure 3.2. PL Continue example

2. **Trace Exercise:** requires the tested subject to trace the drawing pattern of the reference image with the stylus. This is the task that requires only the motion execution phase without complex planning.

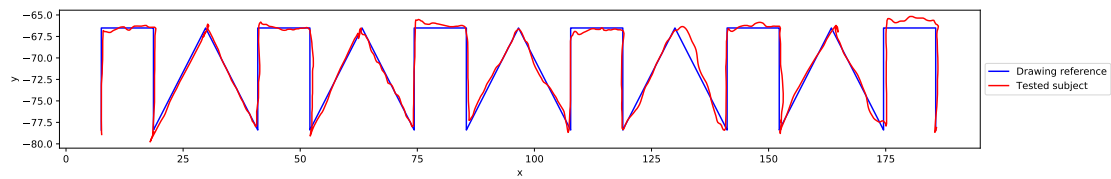


Figure 3.3. PL Trace example





## Chapter 4

# Experimental settings

### 4.1 Hardware

Apple iPad Pro 9.7 (2016 model) and Apple Pencil used for drawing data acquisition. The iPad Pro sample rate with the Apple Pencil is 240 points per second.

### 4.2 Software

To collect the testing data, a special application previously developed in our research group was used, which is named MeDiag Medical Diagnosis. This custom iOS application was developed with the Swift programming language and Xcode IDE (integrated development environment). The application saves acquired data in JSON format and sends them to a remote back-end service. MeDiag provides the practitioner with the possibility to choose from a battery of different tests, and test the subject affected by PD.

#### 4.2.1 Data description

The data format is JSON. Each JSON file has a specific name and collection of fields, filled with meta-information about corresponding drawing, such as  $\{patientId, session, time, type\}$  and drawing data, which is a list of points. Every point is a vector of  $\{x, y, t, p, a, l\}$ , where:

- $x, y$  - coordinates,

- $t$  - time stamp,
- $p$  - pressure,
- $a, l$  - angles of altitude and latitude.

A sample of a JSON file of analysed data is represented below:

```
{
  "data": [
    [
      {
        "p": 0.333333,
        "l": 1.096309,
        "y": 174.293,
        "x": 155.9844,
        "a": 1.112797,
        "t": 577968966.922705
      },
      {
        ...
      }
    ]
  ],
  "hand": "M",
  "time": "2019-04-15 10:56:06 +0000",
  "patientId": "subject_XX",
  "type": "lines"
}
```

### 4.2.2 Data acquisition

The data had been collected on a weekly and daily basis. The weekly collection counts  $\approx 20$  subjects between 18 and 30 years old and covers 8 weeks. The daily collection counts  $\approx 10$  subjects between 18 and 40 years old and covers 5 days. Every session the subject was requested to sit in a comfortable position and to complete four exercises. The following table shows how the values of every test were stored in matrix form:

Table 4.1. Test values table

t	x	y	p	a	l
577968966.922705	155.9844	174.293	0.333333	1.112797	1.096309
...	...	...	...	...	...

## 4.3 Research and development

Research, analysis and the algorithm was performed using the Python programming language and Spyder IDE, both included in the Anaconda Suite that is a free and open-source distribution of Python, that aims to simplify package management and deployment [32]. The following open-source Python libraries were used to perform this research:

- Json, NumPy and Pandas: for data reading, storage and manipulation
- Matplotlib: for figure plotting
- SciPy: for feature generation and statistical analysis
- Descartes: for feature generation
- Scikit-learn: for training and validation of linear regression models
- Statistics: to compute the standard deviation
- Shapely: for feature generation
- Statsmodels: for models building

The special software developed to process the acquired data for the present research is capable of:

- Data extraction from JSON files

The extracted data are encapsulated in a Pandas DataFrame

- Features Extraction

From the collected data a sequence of operations are made to obtain numeric features, for more details refer to section [5.1](#)

- Statistical analysis on features

Different statistical tests are made on the extracted features, for more details refer to sections [4.4.1](#) and [6.1](#)

- Feature filtering

On the base of statistical tests results a ranking of the features is done

- Model building

Backward feature elimination to build the model is applied, for more details refer to sections [4.4.2](#) and [6.2](#)

- Plotting of all the obtained results

To favourite a wide view of the results plots are produced, for more details refer to chapter [7](#)

## 4.4 Mathematical method

This section provides a brief overview of the knowledge required to understand the workflow of the thesis. It is not intended to provide a complete summary of the used mathematical methods. To fully understand these concepts please refer to the cited books.

### 4.4.1 Statistical one population sample test

The one-sample t-test or one sample population test is used to compare a single sample mean to a specified constant hypothesized to be the population mean [\[35\]](#). In the present research, the value used for the constant is zero

which will indicate whether any change occurred between the two instant of time points for the original measured features, or else the population at the beginning and the end of the experiment. The one-sample t-test is a parametric test that should be used when you have a small sample of the whole population. If the mean change score is not significantly different from zero, no significant change occurred. The smaller is the p-value more significant is the change occurred.

The following analysis were conducted using the package `scipy.stats.ttest_1samp()` [34].

### 4.4.2 Linear regression

Linear regression is a linear method to modelling the relationship between dependent variable and one or more independent variables (explanatory variables). In linear regression, the relationships are modelled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. The case of one explanatory variable is called simple linear regression. [36].

The model of linear regression is represented by the following equation:

$$y_i = \beta_0 + \beta_{i,1}x_{i,1} + \beta_{i,2}x_{i,2} + \dots + \beta_{i,n}x_{i,n} + \epsilon$$

Where:

- $y_i$  variable called the regressand, endogenous variable, response variable, measured variable, criterion variable, or dependent variable.
- $\beta_0$  is the intercept term.
- $x_i$  are fixed values chosen prior to observing the dependent variable.
- $\epsilon$  is called the error term, disturbance term, or sometimes noise (in contrast with the "signal" provided by the rest of the model).

To reduce the error  $\epsilon$  linear regression models are often fitted using the least squares approach. The equation that we want to minimize is:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Where:

- $y_i$  is the true value of the observation i.

- $\hat{y}_i$  is the estimated value of the observation  $i$ .

So we should minimize the following term:

$$\sum_{i=1}^n (y_i - (\beta_0 + \beta_{i,1}x_{i,1} + \dots + \beta_{i,n}x_{i,n} + \epsilon))^2$$

solving the system to find  $\beta_0, \dots, \beta_n$ , called regression coefficients.

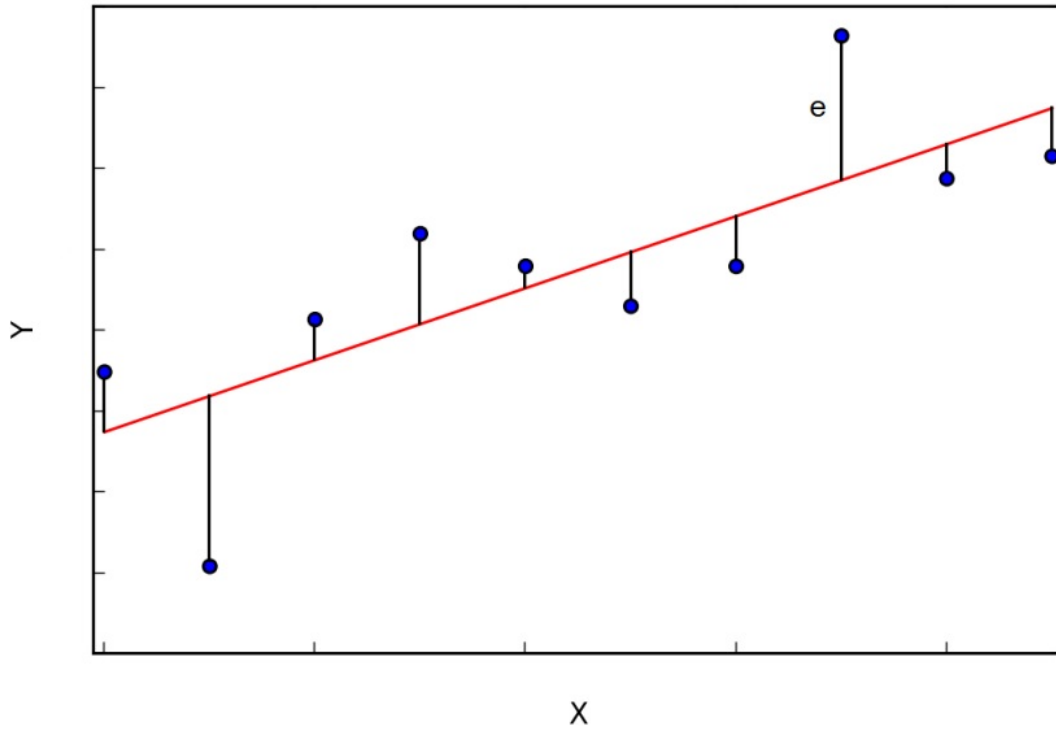


Figure 4.1. Least squares error approach

The following analysis were conducted using the package *statsmodels* [37].

#### 4.4.3 Cross-validation

Cross-validation is one of many model validation techniques for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. In a prediction problem, a model is usually given a dataset of known data on

which training is run (training dataset), and a dataset of unknown data (or first seen data) against which the model is tested (called the validation dataset or testing set). The goal of cross-validation is to test the model's ability to predict new data that was not used in estimating it, in order to flag problems like overfitting or selection bias and to give an insight on how the model will generalize to an independent dataset (i.e., an unknown dataset, for instance from a real problem).

One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set). To reduce variability, in most methods multiple rounds of cross-validation are performed using different partitions, and the validation results are combined (e.g. averaged) over the rounds to give an estimate of the model's predictive performance.

In summary, cross-validation combines (averages) measures of fitness in prediction to derive a more accurate estimate of model prediction performance [39].

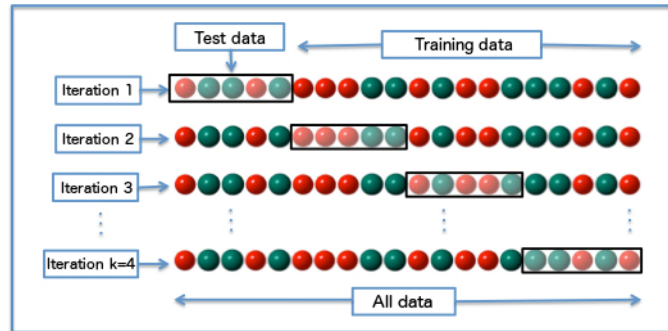


Figure 4.2. Diagram of k-fold cross-validation with  $k=4$ .

The following analysis were conducted using the package `sklearn.model_selection.cross_val_score` [38].





# Chapter 5

## Features

### 5.1 Feature extraction

Feature extraction starts from an initial set of measured data and builds, by the application of functions, derived values (features) intended to be informative and non-redundant. Methods can be divided into three categories:

1. **Unary functions**: methods, with one argument as input, generate a single feature from a single feature. Sample unary functions would be trigonometric, square, root, exponentiation, etc...
2. **Binary functions**: methods, with two arguments as input, generate a single feature from two features. Sample binary functions would be all arithmetic functions addition, multiplication, etc...
3. **Array functions**: methods, with multiple arguments as input, generate a single feature from an array of features. For example statistical functions max, min, median, etc...

In the present research each data point, itself was an array of  $\{x, y, t, p, l, a\}$  scalar values. Any array-like feature of the Drawing object is the result of a set of points converted into a feature by using statistical functions such as  $\{mean, median, min, max, standard\ deviation, ...\}$ . From  $\{x, y\}$  coordinates of data points and time  $t$  was possible to produce arrays of kinematic features and geometric, such as length, duration, velocity, acceleration, jerk, slope and similarly apply statistical functions to them to generate unique higher-order features.

An example can help understanding the process. Starting from the collected data points:

$$\vec{X} = [x_1, x_2, \dots, x_n]$$

$$\vec{Y} = [y_1, y_2, \dots, y_n]$$

$$\vec{T} = [t_1, t_2, \dots, t_n]$$

An array is fulfilled with velocity between each data point:

$$\vec{v} = \lim_{\Delta t \rightarrow 0} \frac{\Delta \vec{r}}{\Delta t}$$

$$\vec{v}_n = \frac{\sqrt{(x_n - x_{n-1})^2 + (y_n - y_{n-1})^2}}{t_n - t_{n-1}}$$

$$\vec{V} = [v_1, v_2, \dots, v_n]$$

The feature *velocity\_mass* is computed as:

$$velocity\_mass = \sum_{i=1}^n |v_i| \quad \text{with} \quad -\infty < v_i < +\infty$$

and the feature *velocity\_mean* is computed as:

$$velocity\_mean = \frac{1}{n} \sum_{i=1}^n |v_i| \quad \text{with} \quad -\infty < v_i < +\infty$$

## 5.2 Feature categories

In Drotar et al. (2016) [31] a large variety of features describing movements of the pen tip is suggested. These features include kinematic parameters and parameters describing contact of the pen tip with the screen. Among kinematic parameters are average values of velocities, accelerations, jerks etc. Also parameters describing directional changes and their number are included. Contact of the pen with the screen is described by applied pressure, number and temporal characteristics of the strokes. In addition to the parameters available from the literature, motion mass parameters adopted for the fine motor case in Nõmm et al. [3, 5] is evaluated in this thesis. The formal definition of the parameters used in this research is presented below.

### 5.2.1 Pressure, geometric and kinematic features

Kinematic and pressure characteristics of handwriting are used in various recent research papers of Nömm et al. [3], Zham et al. [27], San Luciano et al. [26]. Those characteristics have proven a high level of discrimination power in the analysis of PD subjects. Kinematic and pressure feature are vectors of points [p1, p2, ..., pn]. Velocity, acceleration, jerk, slope and pressure higher-order features are generated through the application of statistical functions. Along with standard [median, mean, mass] features, two higher-order features are proposed:

- **Number of changes** – of a specific pressure or kinematic feature. Standard function *argrelextrema()* of Scipy library was applied to evaluate the number of changes. *argrelextrema()* is a sliding-window based method and calculate local extrema of the arbitrary array. In this study, the number of extrema points are used as the number of changes of a particular kinematic or pressure feature.
- **Penalized parameters** – Many subjects slowed down their motions intentionally to gain more precision in tracing the lines. In order to “punish” such activity corresponding value of the motion mass parameters per second will be used.

### 5.2.2 Drawing features

Drawing-related features are defined and evaluated in the context of the corresponding Drawing entity. In our case, Drawing is the whole Luria pattern or Spiral, represented by a vector of points [p1, p2, ..., pn] where [p1, pn] are starting and ending points of the drawing. Along with standard [trajectory\_length, duration] features, one higher-order feature was proposed only for PL Trace exercise:

- **Area difference** – between the reference image and the subject drawn, see Figure 5.1.

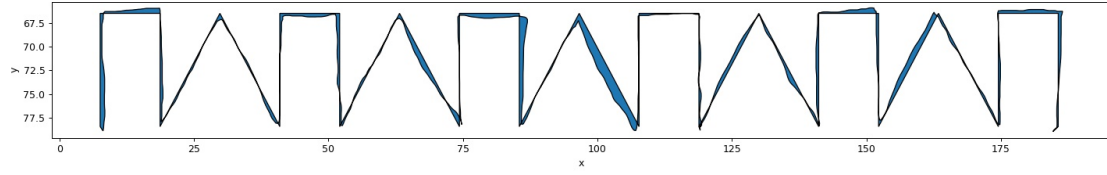


Figure 5.1. Area Difference between PLtrace and reference\_PLtrace

### 5.3 Feature naming

Identification name of the feature is generated by concatenating at least two out of three main components, as explained in the following table showing the possible combinations of functions to produce different features:

Table 5.1. Feature naming table

Prefix	Feature	Suffix	High-order feature
diff_	velocity	_mass	_mean
	acceleration	_diff	_nc (number of changes)
	pressure		_penalized
	slope		_std
	jerk		
	azimuth		
	altitude		

Every element of the feature column must join at least a component of the suffix column. When a high-order feature is generated from a motion mass feature *\_mass* sometimes is omitted to keep the name short, e.g. *velocity\_mean* instead of *velocity\_mass\_mean*. The outcome of the previous step can join one high-order feature and, optionally, the prefix.

Notice that *\_penalized* can not apply to azimuth and altitude and can be applied only to motion mass features, e.g. *velocity\_mass\_penalized*. Is not possible to generate a feature only with a prefix and an element of the feature column, e.g. *diff\_velocity*.

Some examples are given for better comprehension: *velocity\_mean*, which

means the mean computed across all the velocity computed between each adjacent point, *diff\_velocity\_mean*, which means the difference between each adjacent value of the feature *velocity\_mean*, or *velocity\_mass*, which means the absolute value of the summation of all velocity, etc...

### 5.3.1 Feature details

Starting from the collected data points, motion mass features are computed:

- *velocity\_mass* velocities for each collected points and sum of their absolute value is computed

$$velocity\_mass = \sum_{i=1}^n |v_i| \quad \text{with} \quad -\infty < v_i < +\infty$$

- *acceleration\_mass* accelerations for each collected points and sum of their absolute value is computed

$$acceleration\_mass = \sum_{i=1}^n |a_i| \quad \text{with} \quad -\infty < a_i < +\infty$$

- *pressure\_mass* the sum of the absolute value of each collected point is computed

$$pressure\_mass = \sum_{i=1}^n |p_i| \quad \text{with} \quad -\infty < p_i < +\infty$$

- *slope\_mass* slopes for each collected points and sum of their absolute value is computed

$$slope\_mass = \sum_{i=1}^n |s_i| \quad \text{with} \quad -\infty < s_i < +\infty$$

- *jerk\_mass* jerks for each collected points and sum of their absolute value is computed

$$jerk\_mass = \sum_{i=1}^n |j_i| \quad \text{with} \quad -\infty < j_i < +\infty$$

- *azimuth\_mass* the sum of the absolute value of each collected point is computed

$$azimuth\_mass = \sum_{i=1}^n |a_i| \quad \text{with} \quad -\infty < a_i < +\infty$$

- *altitude\_mass* the sum of the absolute value of each collected point is computed

$$altitude\_mass = \sum_{i=1}^n |a_i| \quad \text{with} \quad -\infty < a_i < +\infty$$

It is possible to do all the previous calculation on the differences between each adjacent points, this is noted with *\_diff* instead of *\_mass*.

To obtain any other high-order feature is sufficient to apply one of the previously mentioned functions (see section 5.1) to motion mass features described, such as:

- mean

$$feature\_mean = \frac{1}{n} \sum_{i=1}^n |f_i| \quad \text{with} \quad -\infty < f_i < +\infty$$

- standard deviation

$$feature\_std = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - \bar{f})^2}$$

- number of changes

$$argrelextrema(\vec{F})$$

- penalization

$$feature\_mass\_penalized = \frac{feature\_mass}{t}$$

- difference

$$diff\_feature = f_i - f_{i-1} \quad \text{where } i \text{ is a test session}$$

## 5.4 Feature List

For PL Trace a total of 47 numeric feature are computed:

<i>trajectory_length</i>	<i>pressure_mass_penalized</i>
<i>duration</i>	<i>slope_mass_penalized</i>
<i>velocity_mass</i>	<i>jerk_mass_penalized</i>
<i>acceleration_mass</i>	<i>velocity_std</i>
<i>pressure_mass</i>	<i>acceleration_std</i>
<i>slope_mass</i>	<i>pressure_std</i>
<i>jerk_mass</i>	<i>slope_std</i>
<i>diff_pressure_mass</i>	<i>jerk_std</i>
<i>diff_slope_mass</i>	<i>diff_pressure_std</i>
<i>diff_jerk_mass</i>	<i>diff_slope_std</i>
<i>velocity_mean</i>	<i>azimuth_mass</i>
<i>acceleration_mean</i>	<i>altitude_mass</i>
<i>slope_mean</i>	<i>azimuth_mean</i>
<i>slope_diff_mean</i>	<i>altitude_mean</i>
<i>pressure_mean</i>	<i>azimuth_nc</i>
<i>pressure_diff_mean</i>	<i>altitude_nc</i>
<i>jerk_mean</i>	<i>azimuth_std</i>
<i>jerk_diff_mean</i>	<i>altitude_std</i>
<i>pressure_nc</i>	<i>diff_azimuth_nc</i>
<i>velocity_nc</i>	<i>diff_altitude_nc</i>
<i>acceleration_nc</i>	<i>azimuth_diff_mean</i>
<i>jerk_nc</i>	<i>altitude_diff_mean</i>
<i>velocity_mass_penalized</i>	<i>area_difference</i>
<i>acceleration_mass_penalized</i>	

For PL Continue all the previous feature except *area\_difference* are computed, 46 in total.





## Chapter 6

# Formal problem statement and Methodology

Recent studies determine that Luria's alternating series is capable of highlighting the change in motion and smoothness features of neurology disease affected subjects [3]. However adaptation to those exercises has not been properly studied. The research goal is to determine the presence, if any, of adaptation. In order to achieve this goal, the relations between the categories of the features will be explored. Where it is intended to investigate only relations between the categories of the features that are not linked in an obvious way as an e.g. Area difference and velocity.

The research questions are:

1. Are the proposed experimental settings sensitive enough to study adaptation on fine motor motions?
2. "Filter" features.
  - Find the features, if any, that changes the most.
3. Find models, if any, that explain the relationship between features.

In order to answer the research questions, the following was done:

1. Question number one:

Statistical hypothesis testing.

2. Question number two:

Statistical hypothesis testing to see which features changes significantly.

3. Question number three:

On the base of the previous steps model building with Back Feature Elimination technique is applied.

## **6.1 Statistical analysis**

To answer question one statistical hypothesis testing was used. The main goals were:

- Discriminate between the beginning and the ending part of the same test.
- Discriminate between non-trained and trained subject repeating the same test.

### **6.1.1 Statistical analysis on test exercise performed in different sessions**

In order to verify if a significant change occurred in the whole sample for every feature, the difference between the value of the selected feature and the value of the same feature in the next session was computed. Those differences (one for each subject  $S$ ) were put in an array on which one population t-test was performed. The p-value highlights which ones are the features that changed the most between the two sessions of the test. Refer to section [4.4.1](#) for more details. The difference can be computed in two different ways:

1. Between the first and the second value of the feature, the second and the third value of the feature etc.
2. Between the first value of the feature and all the others

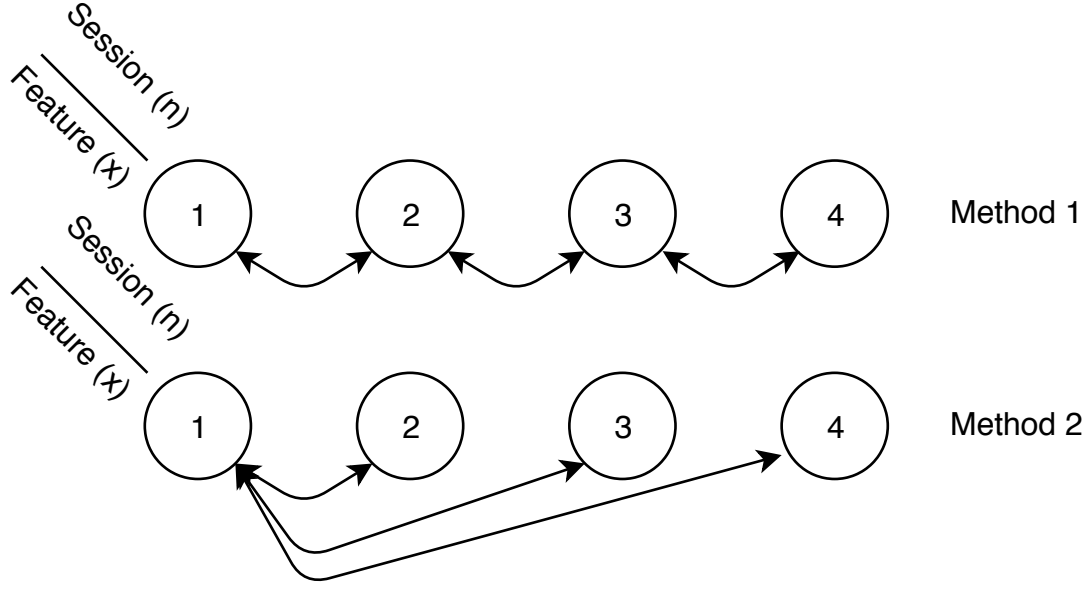


Figure 6.1. Method to compute differences between features

The data collected on a weekly and daily basis were used separately to run one population t-test. The test is performed on the differences between two values of a single feature in two different sessions. To compute those differences the first or last two sessions are used:

$$\Delta(1,2) = F_x(1) - F_x(2) \quad \text{first two sessions}$$

$$\Delta(n-1, n) = F_x(n-1) - F_x(n) \quad \text{last two sessions}$$

The comparison between the differences of the first two sessions highlights a significant change, if any, between two consecutive tests. Instead in the comparison of differences of the first session with the last one will describe a significant change in two non-consecutive tests, which means that the change occurs across or after a certain number of tests.

Choosing one possibility between  $\Delta(1,2)$  and  $\Delta(n-1, n)$ , the differences computed for every subject  $\Delta_{S_i}$  are stored in array form:

$$\vec{\Delta} = [\Delta_{S1}, \Delta_{S2}, \Delta_{S3}, \dots, \Delta_{Sn}] \quad \text{where e.g. } \Delta_{S1} = \Delta(1,2) \quad \text{for subject S1}$$

Then p-value is computed:

$$\text{p-value} = \text{ttest}(\vec{\Delta})$$

If the research settings are sensitive enough for our research scope, application of statistical analysis is expected to provide the following:

- p-values that will describe the discrimination power of features
- Ranking of features

The following numerical example go through the main parts of the process, the total number of tests collected for each subject is equal to five:

- choosing the subject S1, one feature was selected (e.g *acceleration\_nc*) and computed for each test,

$$\text{acceleration\_nc} = [64, 60, 53, 65, 54]$$

- the first method in the Figure 6.1 was applied and the differences computed,

$$\text{acceleration\_nc\_diff\_between\_tests} = [4, 7, -12, 11]$$

- supposed that difference between first and second session was preferred  $\Delta(1,2)$ ,

$$\Delta_{S1} = 64 - 60 = 4$$

- the previous steps are repeated for each subjects. At the end  $\vec{\Delta}$  will be fulfilled of differences computed on every subject:

$$\vec{\Delta} = [4, 7, 8, 15, 16, -24, 0, -1]$$

- then, p-value is computed:

$$\text{p-value} = \text{ttest}(\vec{\Delta}) = 0.505$$

In this specific case the p-value, with a significance level of  $\alpha = 0.05$ , means that there was no significant change in the feature *acceleration\_nc* for the population sample between the first and second sessions.

### 6.1.2 Statistical analysis on individual test exercise

The next step was searching for adaptation within the same test. To answer this question the following procedure was implemented. The new procedure consists of the splitting of the PL exercise into  $n$  intervals:

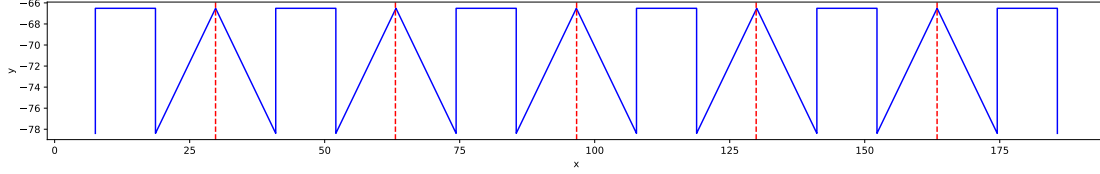


Figure 6.2. PL series divided in  $n$  intervals,  $n=6$

In this case, the difference between the values of a feature is computed separately for each interval. One population test is performed for every interval and the computed p-values are used to plot graphs describing the evolution of the significance level of a feature in different intervals. In the case of intervals, choosing one possibility between  $\Delta(1,2)$  and  $\Delta(n-1, n)$ , the differences computed can be also cumulative:

$$\Delta_{cum1} = \Delta_1 \quad \text{where} \quad \Delta_1 \quad \text{means the difference in the first interval}$$

$$\Delta_{cum2} = \Delta_{cum1} + \Delta_2$$

$$\Delta_{cum3} = \Delta_{cum2} + \Delta_3$$

...

$$\Delta_{cum(n)} = \Delta_{cum(n-1)} + \Delta_n$$

The cumulative differences computed for every subject are stored in different arrays, one for each interval:

$$\Delta_{int1}^{\rightarrow} = [\Delta_{cumS1}, \Delta_{cumS2}, \dots, \Delta_{cumSn}]$$

where e.g.  $\Delta_{cumS1} = \Delta_{cum(n)}$  for the subject S1 in the first interval

Then p-value is computed:

$$\text{p-value}_{int1} = \text{ttest}(\Delta_{int1}^{\rightarrow})$$

the process is iterated for every subject and the various p-values are used to plot graphs describing the evolution of the significance level of a feature in different intervals.

Application of statistical analyse is expected to provide the following:

- p-values that will describe the discrimination power of features for each interval
- Ranking of features for each interval

The following numerical example go trough the main parts of the process with *number of intervals* setted to two, the total number of tests collected for each subject is equal to five:

- choosing the subject S1, one feature was selected (e.g *acceleration\_nc*) and computed for each test,

$$acceleration\_nc = [47, 49, 48, 39, 43] \quad \text{for the 1st interval}$$

$$acceleration\_nc = [43, 42, 40, 46, 42] \quad \text{for the 2nd interval}$$

- the first method in the Figure 6.1 was applied and the differences computed,

$$acceleration\_nc\_diff\_between\_tests = [-2, 1, 9, -4] \quad \text{for the 1st interval}$$

$$acceleration\_nc\_diff\_between\_tests = [1, 2, -6, 4] \quad \text{for the 2nd interval}$$

- supposed that difference between the last two sessions was preferred  $\Delta(n-1, n)$ ,

$$\Delta_{cum1} = \Delta(n-1, n)_{int1} = 39 - 43 = -4 \quad \text{for the 1st interval}$$

$$\Delta_{cum2} = \Delta_{cum1} + \Delta(n-1, n)_{int2} = -4 + (46 - 42) = 0 \quad \text{for the 2nd interval}$$

- At the end  $\vec{\Delta}_{int1}$  and  $\vec{\Delta}_{int2}$  will be fulfilled of differences:

$$\vec{\Delta}_{int1} = [\Delta_{cumS1}, \Delta_{cumS2}, \dots, \Delta_{cumSn}]$$

where e.g.  $\Delta_{cumS1} = \Delta_{cum(n)} = -4$  for the subject S1 in the 1st interval end,

$$\vec{\Delta}_{int2} = [\Delta_{cumS1}, \Delta_{cumS2}, \dots, \Delta_{cumSn}]$$

where e.g.  $\Delta_{cumS1} = \Delta_{cum(n)} = 0$  for the subject S1 in the 2nd interval

- the previous steps are repeated for each subjects and interval:

$$\Delta_{int2}^{\rightarrow} = [-4, -7, 1, -3, 7, 9, -5, 1]$$

$$\Delta_{int2}^{\rightarrow} = [0, -10, 3, -5, 4, 10, -10, 8]$$

- then, p-value is computed for each interval:

$$\text{p-value}_{int1} = \text{ttest}(\Delta_{int1}^{\rightarrow}) = 0.952$$

$$\text{p-value}_{int2} = \text{ttest}(\Delta_{int2}^{\rightarrow}) = 0.952$$

In this specific case the p-values, with a significance level of  $\alpha = 0.05$ , means that there was no significant change in the feature *acceleration\_nc* for the population sample in the two intervals.

## 6.2 Model building

The next step was to analyse all the data searching for a suitable linear regression model. The procedure consists of different phases:

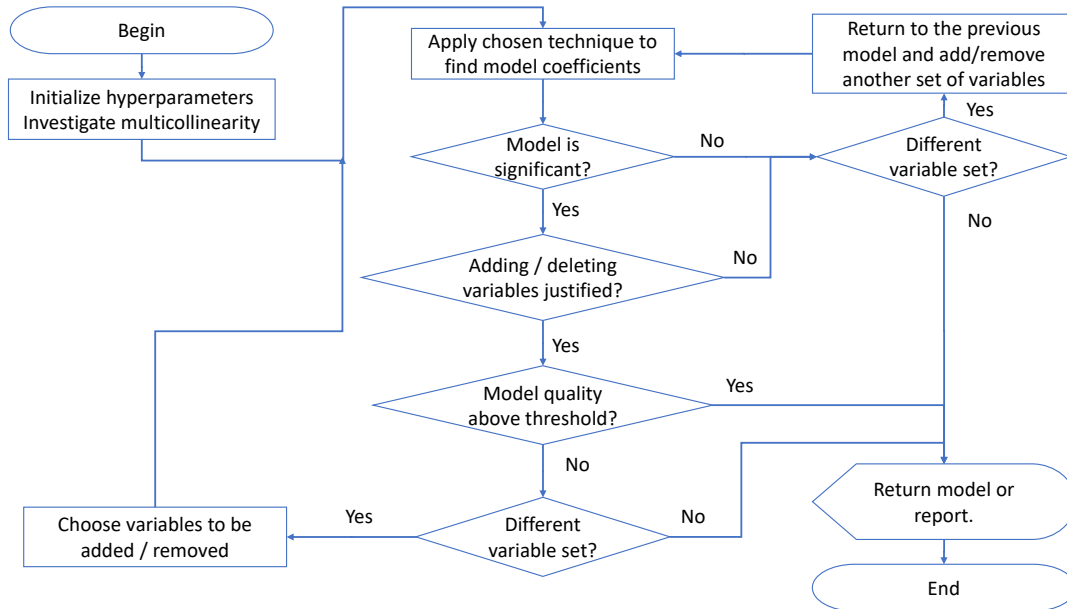


Figure 6.3. Linear model building process [33]

1. Hyper-parameters are chosen or determined
  - e.g. Forward feature selection/Backward feature elimination/Bidirectional feature elimination, correlation threshold and level of significance
2. Apply one technique to build the model
3. Evaluate the quality (error, e.g. mean square) and significance (F-test, t-test) of the model
4. If criteria are not met return to the previous model and choose another set of variables to add/delete
  - From the 2nd iterate use F-test to prove if model was improved or not by adding/deleting variables
5. If model does not improve return to the previous model and select another variable to add/delete
6. If criteria are met return the built model
7. Iterate until criteria are met



### 6.2.1 Data preparation

All the data were represented in a matrix form. The columns represent the features and the rows the observation points for every subject.

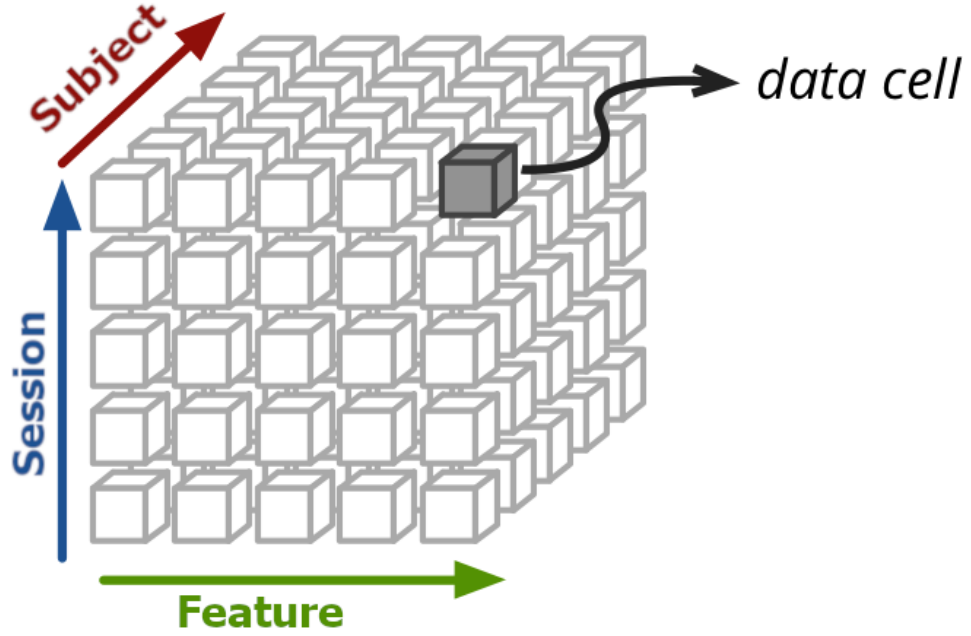


Figure 6.4. Matrix representation of the observation used for LM building

Before starting to build the models the data set is divided into `train_set` and `test_set` respectively 70% and 30%. The linear regression models computed are static, no different instant of time are considered:

$$y_i = \beta_0 + \beta_{i,1}x_{i,1} + \beta_{i,2}x_{i,2} + \dots + \beta_{i,n}x_{i,n} + \epsilon$$

Refer to section 4.4.2 for more details.

### 6.2.2 Backward feature elimination

There are three widely used techniques for linear model building:

- Forward feature selection - build the model by adding features as predictors and evaluating the accuracy of the model at each step.

- Backward feature elimination - build the model with all features as predictors and at each step remove the features that not improve the accuracy of the model.
- Bidirectional feature elimination - a combination of the above, testing at each step for variables to be included or excluded.

The one chosen for this research is Backward feature elimination:

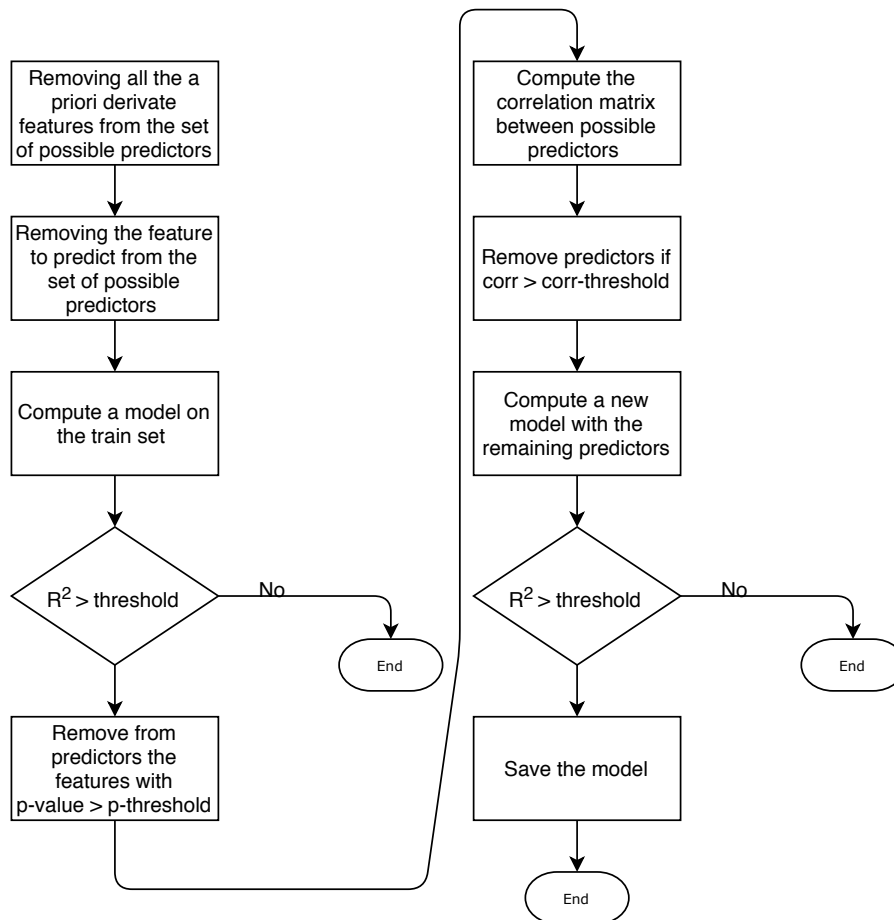


Figure 6.5. Backward feature elimination process

1. Removing all the a priori derivative features from the set of possible predictors

e.g. If the current feature is *acceleration\_mean* all features related to the same measurement will be removed, like *acceleration\_mass*.

2. Removing the feature to predict from the set of possible predictors
3. Compute a model on the training set
4. Evaluate model  $R^2$  Adjusted against a threshold  
 $R^2$  Adjusted explain how much of the total variation of the data can be explained by the model.
5. Remove from the set of possible predictors all the features that have a p-value greater than a given p-threshold
6. Compute the correlation matrix between the set of possible predictors
7. Remove Highly Correlated predictors from the set of possible predictors  
Predictors are Highly Correlated if greater then a given corr-threshold
8. Compute a new model with the remaining features in the set of possible predictors
9. Evaluate model  $R^2$  Adjusted against a threshold
10. Save the model

The whole process is repeated two times to recompute p-value and check if the significance of the model is still acceptable with the remained set of possible predictors.

### 6.2.3 Model validation

After the model is built it has to be validated. K-fold cross-validation was used with  $k=10$ . Refer to section 4.4.3 for more details. If the model reaches the validation criteria is saved.

Application of model building technique is expected to provide a model, if any, that describes the changing occurred on all the sessions.

#### 6.2.4 Modelling individual test exercise

The same method used to build the model for the whole test is used with the technique that consists in the subdivision of the whole test in a given number of intervals. For each part, all the feature extraction is performed in the same way. Each interval is treated as a separate observation of a subject, thanks to this approach the absolute value of the accumulated error is reduced and the number of observation to compute the model grows:

$$n_{obs} = n_{subject} * n_{sessions} * n_{intervals}$$

e.g. on the weekly collection  $13.600 = 17 * 8 * 100$

To establish the proper amount of intervals the sequent approach was used:

1. Compute R2 adj. for the model on the given amount of intervals
2. Increment the number of intervals

$$n_{intervals} = n_{intervals} + 10 \text{ (or 5)}$$

3. Iterate

Application of model building technique is expected to provide a model, if any, that describe the relationships between features of each interval for all sessions.

# Chapter 7

## Result

In the following pages, the results obtained with the previously explained methodologies will be discussed and explained. Only the most valuable results will be included in this chapter. All research questions will be answered tracing the same structure of the previous chapter.

*Notice* that from this chapter the terms **weekly** and **daily** collection will refer to the set of features computed for each test on initial data, refer to sections 5.1 and 5.4, and not any more to the initial collected data, refer to section 4.2.2. To avoid misunderstandings, this difference is emphasized with the two terms bolded.

### 7.1 Statistical analysis

#### 7.1.1 Statistical analysis on test exercise performed in different sessions

One population test results on the **weekly** based collection, with a significance level of  $\alpha = 0.05$ , has demonstrated that:

1. For the Trace exercise described by 47 numeric features

The following 3 (see Table 7.1) differ significantly between the first and the second session. The result mean that there is a change for the whole sample of subjects in those 3 features between the first and second session.

Table 7.1. Features — STATISTICAL ANALYSIS — PLtrace — Weekly — first and second session

Feature	t-statistic	p-value
diff_pressure_mass	3.9179	0.0012
pressure_diff_mean	3.4871	0.0030
altitude_diff_mean	2.1869	0.0439

The following 3 (see Table 7.2) differ significantly between the first and the last session. The result mean that there is a change for the whole sample of subjects in those 3 features between the first and last session.

Table 7.2. Features — STATISTICAL ANALYSIS — PLtrace — Weekly — first and last session

Feature	t-statistic	p-value
slope_mean	-2.528	0.0223
slope_mass_penalized	-2.479	0.0246
slope_std	-2.960	0.0092

## 2. For the Continue exercise described by 46 numeric features

The following 1 (see Table 7.3) differ significantly between the first and the second session. The result mean that there is a change for the whole sample of subjects in this feature between the first and second session.

Table 7.3. Features — STATISTICAL ANALYSIS — PLcontinue — Weekly — first and second session

Feature	t-statistic	p-value
diff_jerk_mass	2.1568	0.0465

The following 1 (see Table 7.4) differ significantly between the first and the last session. The result mean that there is a change for the whole sample of subjects in this feature between the first and last session.

Table 7.4. Features — STATISTICAL ANALYSIS — PLcontinue — Weekly — first and last session

Feature	t-statistic	p-value
azimuth_diff_mean	2.392	0.0293

One population test results on the **daily** based collection, with a significance level of  $\alpha = 0.05$ , has demonstrated that:

1. For the Trace exercise described by 47 numeric features

No significant change was observed. The result mean that there is no change for the whole sample of subjects between the first and second or last session.

2. For the Continue exercise described by 46 numeric features

the following 5 differ significantly between the first and the second session. The result mean that there is a change for the whole sample of subjects in those 5 features between the first and second session.

Table 7.5. Features — STATISTICAL ANALYSIS — PLtrace — Daily — first and second session

Feature	t-statistic	p-value
acceleration_std	2.7778	0.0273
jerk_std	2.6866	0.0312
azimuth_mass	-6.117	0.0004
azimuth_mean	-3.754	0.0071
diff_azimuth_nc	2.5541	0.0378

the following 2 differ significantly between the first and the last session. The result mean that there is a change for the whole sample of subjects in those 2 features between the first and last session.

Table 7.6. Features — STATISTICAL ANALYSIS — PLtrace — Daily  
— first and last session

Feature	t-statistic	p-value
altitude_nc	3.0924	0.0175
diff_altitude_nc	2.9407	0.0216

All the results previously reported show that our research settings are sensible to data collection periodicity (**weekly** or **daily**) and time. However, the lack of a common set of features changing between each session for the same test makes, at least on the bases of the collected data, not possible to model this process as a function of time.

### 7.1.2 Statistical analysis on individual test exercise

*Notice* the sequent graphs represent the evolutions of p-values (*y-axis*) obtained for each interval with one population test technique across intervals (*x-axis*), for more details refer to section 6.1.2. The computed p-values are used to plot graphs describing the evolution of the significance level of a feature at different intervals. The red-line represents the significance level of  $\alpha = 0.05$ . Every line under the red-line indicates that the p-value of the corresponding feature is changed significantly between the two adjacent intervals.

With the intervals technique one population test results on the **weekly** based collection, with a significance level of  $\alpha = 0.05$ , has demonstrated that:

1. For the Trace exercise described by 47 numeric features, the results reported in Figure 7.1 are relative to the differences between the first and second session computed with the cumulative method, for more details refer to section 6.1.2. A subset of features  $\{\text{diff\_pressure\_mass}, \text{altitude\_std}, \text{pressure\_std}, \text{altitude\_diff\_mean}, \text{altitude\_mean}, \text{pressure\_diff\_mean}, \text{altitude\_mass}\}$  changed significantly for at least 7 intervals out of 10 was found. The result mean that there are changes for the whole sample of subjects in this set of features in 7 intervals out of 10.



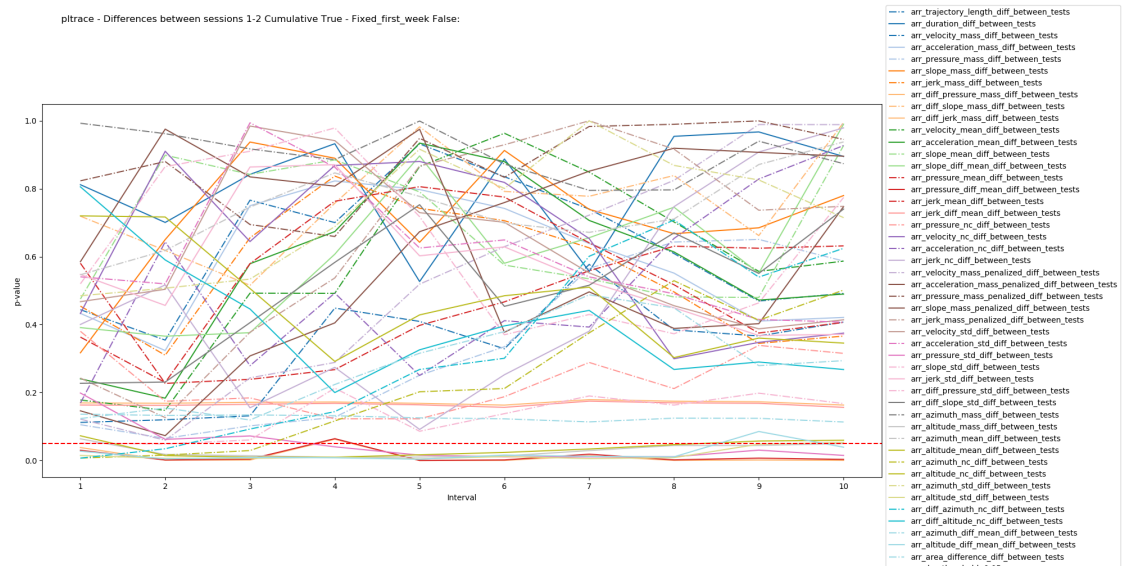


Figure 7.1. PLtrace - Weekly - Differences between sessions 1-2 Cumulative True

- For the Continue exercise described by 46 numeric features, the results reported in Figure 7.2 are relative to the differences between the penultimate and last session computed with the cumulative method, for more details refer to section 6.1.2. A subset of features  $\{altitude\_mean, azimuth\_std, azimuth\_diff\_mean\}$  changed significantly for at least 9 intervals out of 10 was found. The result mean that there are changes for the whole sample of subjects in this set of features in 9 intervals out of 10.

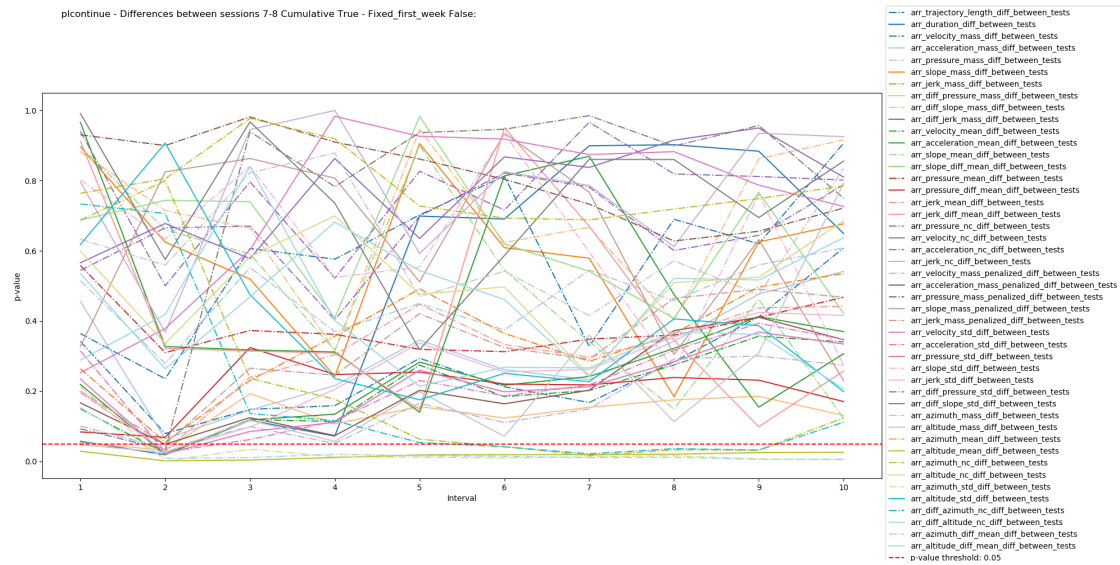


Figure 7.2. PLcontinue - Weekly - Differences between sessions 7-8  
Cumulative True

With the intervals technique one population test results on the **daily** based collection, with a significance level of  $\alpha = 0.05$ , has demonstrated that:

1. For the Trace exercise described by 47 numeric features, the results reported in Figure 7.3 are relative to the differences between the penultimate and last session computed with the cumulative method, for more details refer to section 6.1.2. A subset of features  $\{pressure\_mean, pressure\_mass\_penalized\}$  changed significantly for at least 3 intervals out of 10 was found. The result mean that there are changes for the whole sample of subjects in this set of features in 3 intervals out of 10.

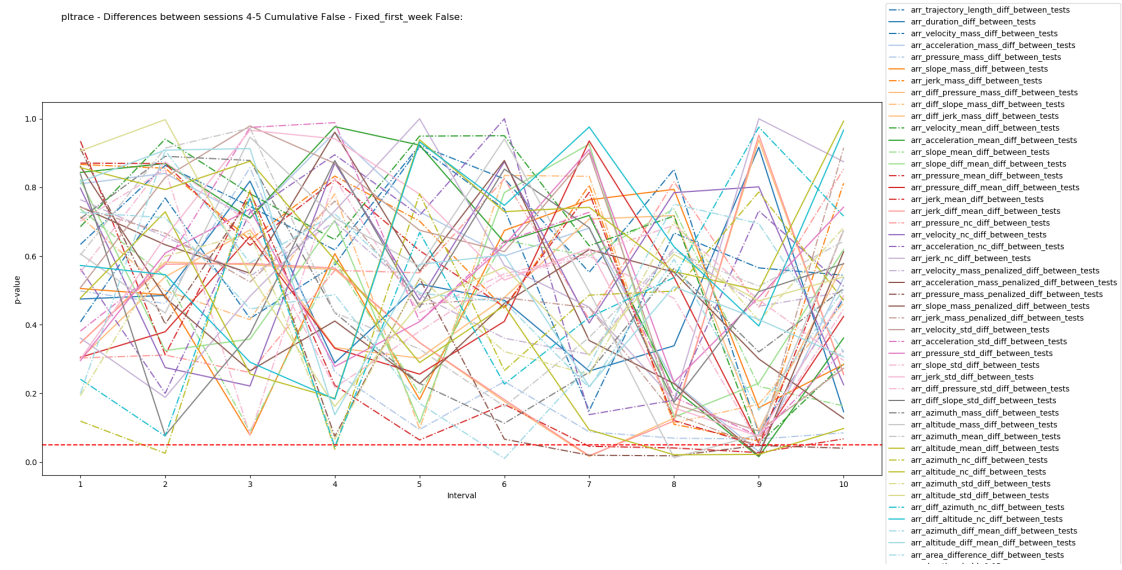


Figure 7.3. PLtrace - Daily - Differences between sessions 4-5 Cumulative False

2. For the Continue exercise described by 46 numeric features, the results reported in Figure 7.4 are relative to the differences between the penultimate and last session computed with the cumulative method, for more details refer to section 6.1.2. A subset of features  $\{trajectory\_length, altitude\_mass, altitude\_nc\}$  changed significantly for at least 3 intervals out of 10 was found. The result mean that there are changes for the whole sample of subjects in this set of features in 3 intervals out of 10.

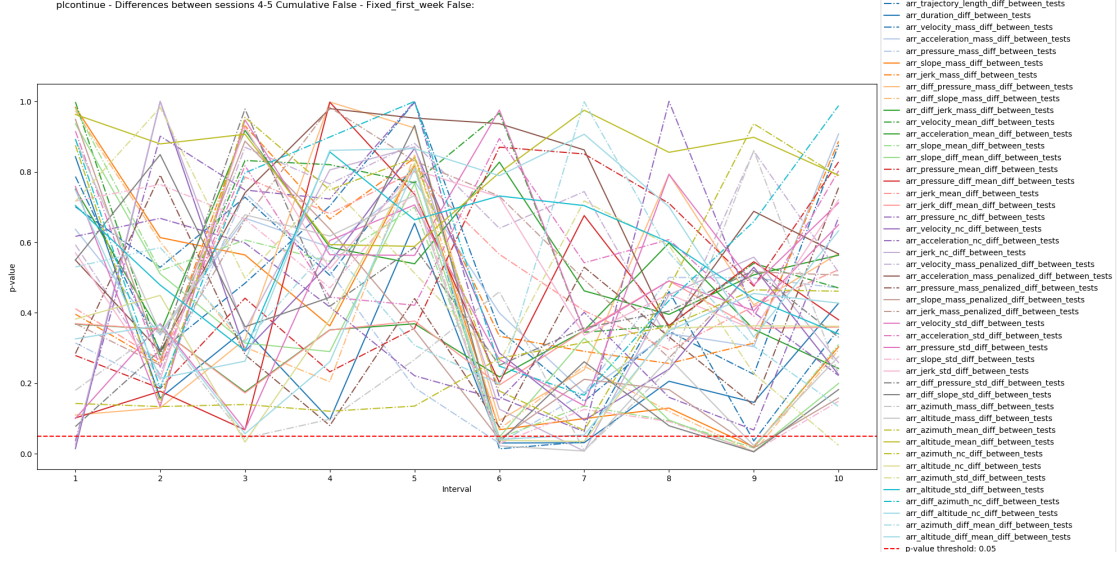


Figure 7.4. PLcontinue - Daily - Differences between sessions 4-5 Cumulative False

All the results previously reported show that changes are occurring within the same test for the whole sample. This means that the subjects are changing during the execution of the task. However, at least on the bases of the collected data, it is not possible to model this process as a function of time. Plus, these results can help to determinate the optimal length of the test needed to evaluate subjects. The length of the test can be adjusted accordingly to the number of intervals where features have a significant change. E.g. if a significant change is observed only in the first three intervals out of ten, there is no need to keep the remaining seven intervals.

## 7.2 Model building

### 7.2.1 Modelling test exercise performed in different sessions

The experimental findings to determine the proper amount of intervals revealing 30 as an acceptable compromise between the performance and the value of  $R^2$  adj, for more details refer to section 6.2.4. The graph in Figure 7.5 show the evolution of the  $R^2$  adj. value  $y$ -axis changing the number of intervals ( $x$ -axis).

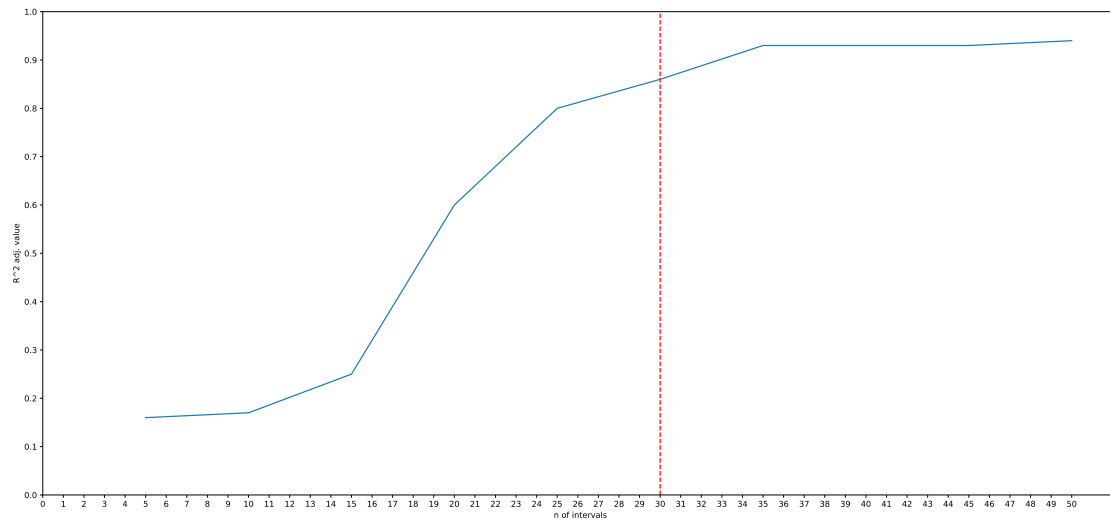


Figure 7.5.  $R^2$  adj. over number of intervals

Tring to model any of the 47 numerical features computed was ineffective. The results, at least on the bases of the collected data, show that no models are capable of explaining adaptation to the tests. The Figure 7.6 shows for every subject the value of the same feature (*y-axis*) over different tests (*x-axis*). The graph clearly shows that some subjects share commons paths highlighted with the same colour. However, there is not a general trend.

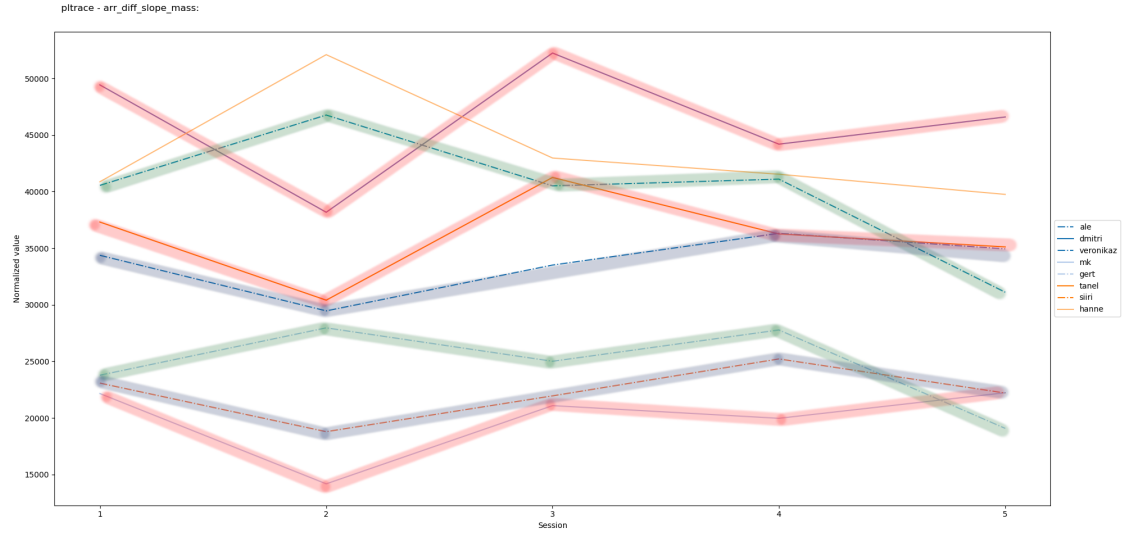


Figure 7.6. Common patterns between subgroup of subjects on the daily collection

These findings show that sub-group of subject share the same pattern. However, no common pattern is found this lead to the necessity of a larger sample in order to find out if sub-groups of the entire population share similar patterns.

### 7.2.2 Modelling individual test exercise

The results, at least on the bases of the collected data, show that there are models capable of explaining the relationship between different features. In particular, the models put in correlation pressure features (P) with the amount of mass (AM), geometric (G), drawing (D) and smoothness features (S). The relationships between pressure features and all the others are investigated because sometimes subjects may unconsciously apply more pressure to gain more support for their hand that may end in terms of performance in better results. The found models can be used as penalization models to mitigate such effects.

The following part show some of the interesting models found for PL Continue on the weekly collection. The first model can predict the number of

changes in acceleration using, as predictors, the number of changes in pressure and the differences between pressure mass. And could be used to penalize subjects that may unconsciously apply more pressure to gain more support for their hand.

**Dep. variable:** acceleration\_nc

**Predictors:** diff\_pressure\_mass, pressure\_nc

**Accuracy (K-fold cross-validation):** 0.93 (+/- 0.17)

**Prediction:**

Table 7.7. OLS Regression Results with acceleration\_nc as dependent variable

<b>Dep. Variable:</b>	acceleration_nc	<b>R-squared:</b>	0.970
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.970
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	4.602e+04
<b>Date:</b>	Fri, 16 Aug 2019	<b>Prob (F-statistic):</b>	0.00
<b>Time:</b>	15:37:01	<b>Log-Likelihood:</b>	-5293.1
<b>No. Observations:</b>	2856	<b>AIC:</b>	1.059e+04
<b>Df Residuals:</b>	2853	<b>BIC:</b>	1.061e+04
<b>Df Model:</b>	2		

	coef	std err	t	P> t	[0.025	0.975]
<b>const</b>	0.2047	0.031	6.529	0.000	0.143	0.266
<b>diff_pressure_mass</b>	0.1928	0.020	9.733	0.000	0.154	0.232
<b>pressure_nc</b>	0.8030	0.004	188.501	0.000	0.795	0.811

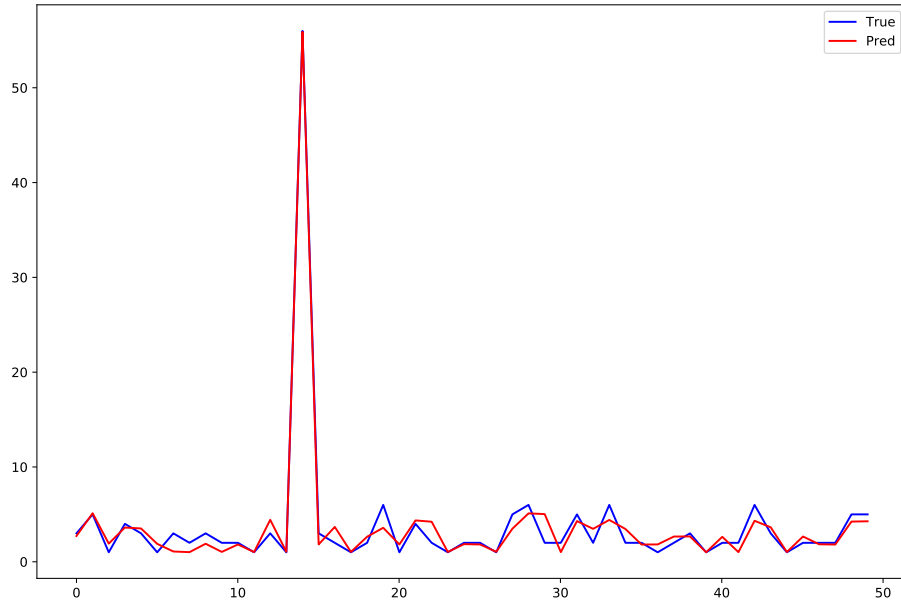


Figure 7.7. Prediction of acceleration\_nc based on pressure features

The second model can predict the number of changes in azimuth angle using, as predictors, the differences between pressure mass, the differences between pressure mean and the number of changes in pressure. And could be used to penalize subjects that may unconsciously apply more pressure to gain more support for their hand.

**Dep. variable:** azimuth\_nc

**Predictors:** diff\_pressure\_mass, pressure\_diff\_mean, pressure\_nc

**Accuracy (K-fold cross-validation):** 0.89 (+/- 0.19)

**Prediction:**



Table 7.8. OLS Regression Results with azimuth\_nc as dependent variable

<b>Dep. Variable:</b>	azimuth_nc	<b>R-squared:</b>	0.938
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.938
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	1.437e+04
<b>Date:</b>	Fri, 16 Aug 2019	<b>Prob (F-statistic):</b>	0.00
<b>Time:</b>	14:24:01	<b>Log-Likelihood:</b>	-12102.
<b>No. Observations:</b>	2856	<b>AIC:</b>	2.421e+04
<b>Df Residuals:</b>	2852	<b>BIC:</b>	2.424e+04
<b>Df Model:</b>	3		

	coef	std err	t	P> t	[0.025	0.975]
<b>const</b>	-0.5320	0.435	-1.224	0.221	-1.385	0.321
<b>diff_pressure_mass</b>	3.8766	0.288	13.445	0.000	3.311	4.442
<b>pressure_diff_mean</b>	-79.3954	16.804	-4.725	0.000	-112.344	-46.446
<b>pressure_nc</b>	5.3298	0.051	105.273	0.000	5.231	5.429

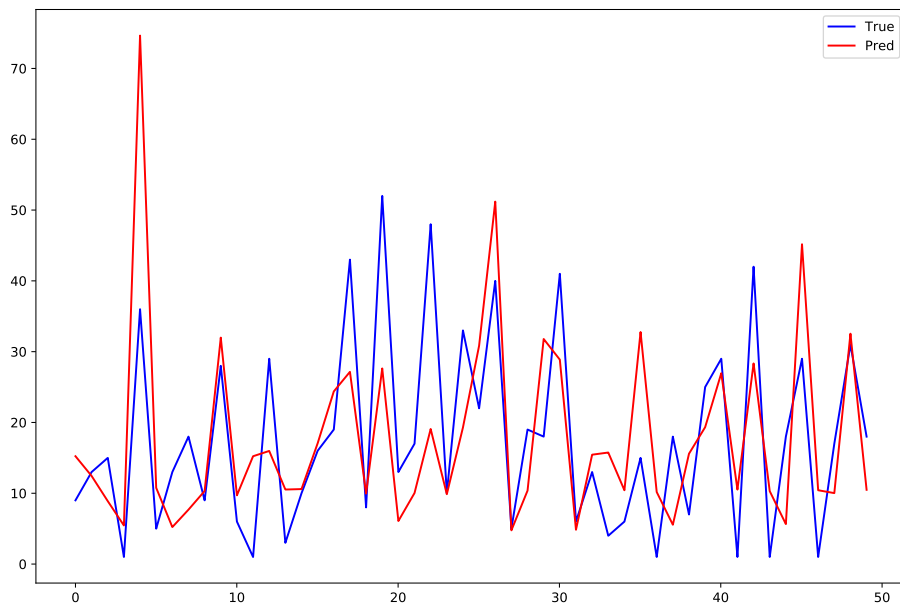


Figure 7.8. Prediction of azimuth\_nc based on pressure features



## Chapter 8

# Conclusion

The main results of the present thesis describe the models computed on the kinematic, geometric, drawing and pressure features allowing to apply corrections to the test result in case of anomalies or cheating by the PD patients. The obtained concept could also be applied in other research fields. The main outcome of the analysis was the linear regression model, capable of explaining the relationships between features, providing average prediction performance around 90%. The proposed model could be included in the decision support framework for Parkinson's disease screening can be potentially adopted by clinicians in medical facilities. Presented methodology can certainly help to obtain more reliable results. The results presented in the present thesis show, that main aims were successfully fulfilled.

The present study can grow in different directions. The most obvious research direction - discover out if sub-groups of the entire population share similar patterns. Without a doubt, it will require a reasonably large group of people.



# Bibliography

- [1] Eusebi, Franchini, De Giorgi, Abraha, Montedori, Casucci, Calabresi, and Tambasco. "Incidence and Prevalence of Parkinson's Disease in the Italian Region of Umbria: A Population-based Study Using Healthcare Administrative Databases." *Neurological Sciences* 40.8: 1709-712. Web.
- [2] Wikipedia contributors. "Parkinson's disease." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 11 Oct. 2019. Web. 14 Oct. 2019.
- [3] Nõmm, Sven, Aaro Toomela, Julia Kozhenkina, and Toomas Toomsoo. "Quantitative Analysis in the Digital Luria's Alternating Series Tests." 2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV) (2016): 1-6. Web.
- [4] Nõmm, Sven, Aaro Toomela, and Jevgeni Borushko. "Alternative Approach to Model Changes of Human Motor Functions." 2013 European Modelling Symposium (2013): 169-74. Web.
- [5] Nõmm, Sven, Aaro Toomela, Martti Vaske, Dan Uvarov, and Pille Taba. "An Alternative Approach to Distinguish Movements of Parkinson Disease Patients." *IFAC PapersOnLine* 49.19 (2016): 272-76. Web.
- [6] Senkiv, Olesja, Sven Nõmm, and Aaro Toomela. "Applicability of Spiral Drawing Test for Mental Fatigue Modelling." *IFAC PapersOnLine* 51.34 (2019): 190-95. Web.
- [7] Esther J Smits, Antti J Tolonen, Luc Cluitmans, Mark Van Gils, Bernard A Conway, Rutger C Zietsma, Klaus L Leenders, and Natasha M Maurits. "Standardized Handwriting to Assess Bradykinesia, Micrographia and Tremor in Parkinson's Disease." *PLoS ONE* 9.5 (2014): E97614. Web.
- [8] Brodaty, Henry, and Cressida M. Moore. "The clock drawing test for dementia of the Alzheimer's type: a comparison of three scoring methods in a memory disorders clinic." *International Journal of Geriatric Psychiatry* 12.6 (1997): 619-27. Web.

- [9] Nõmm, Sven, Konstantin Bardos, Ilja Masarov, Julia Kozhenkina, Aaro Toomela, and Toomas Toomsoo. "Recognition and Analysis of the Contours Drawn during the Poppelreuter's Test." 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA) (2016): 170-75. Web.
- [10] Mari Raudmann, Pille Taba, and Kadri Medijainen. "Handwriting speed and size in individuals with parkinson's disease compared to healthy controls: the possible effect of cueing." *Acta Kinesiologiae Universitatis Tartuensis*, 20:40-47, 2014.
- [11] Wagle Shukla, Aparna, Songthip Ounpraseuth, Michael S Okun, Vickie Gray, John Schwankhaus, and Walter Steven Metzger. "Micrographia and Related Deficits in Parkinson's Disease: A Cross-sectional Study." *BMJ Open* 2.3 (2012): BMJ Open, 2012, Vol.2(3). Web.
- [12] Van Gemmert, Arend W.A, Hans-Leo Teulings, and George E Stelmach. "Parkinsonian Patients Reduce Their Stroke Size with Increased Processing Demands." *Brain and Cognition* 47.3 (2001): 504-12. Web.
- [13] Letanneux, Alban, Jeremy Danna, Jean-Luc Velay, François Viallet, and Serge Pinto. "From Micrographia to Parkinson's Disease Dysgraphia." *Movement Disorders* 29.12 (2014): 1467-475. Web.
- [14] Pinto, Serge, Jean-Luc Velay, and Marieke Longcamp. "Handwriting as a Marker for PD Progression: A Shift in Paradigm." *Neurodegenerative Disease Management* 5.5 (2015): 367-69. Web.
- [15] Sisti, Jonathan A, Brandon Christophe, Audrey Rakovich Seville, Andrew L.A Garton, Vivek P Gupta, Alexander J Bandin, Qiping Yu, and Seth L Pullman. "Computerized Spiral Analysis Using the iPad." *Journal of Neuroscience Methods* 275 (2017): 50-54. Web.
- [16] Ilja Masarov. "Digital clock drawing test implementation and analysis." Master thesis, Tallinn University of Technology, 2017.
- [17] Aghanavesi, Somayeh, Dag Nyholm, Marina Senek, Filip Bergquist, and Mevludin Memedi. "A Smartphone-based System to Quantify Dexterity in Parkinson's Disease Patients." *Informatics in Medicine Unlocked* 9.C (2017): 11-17. Web.
- [18] Mari Raudmann, Pille Taba, and Kadri Medijainen. "Handwriting speed and size in individuals with parkinson's disease compared to healthy controls: the possible effect of cueing." *Acta Kinesiologiae Universitatis Tartuensis*, 20:40-47, 2014.
- [19] Drotar, Peter, Jiri Mekyska, Irena Rektorova, Lucia Masarova, Zdenek

- Smekal, and Marcos Faundez-Zanuy.“Decision Support Framework for Parkinson’s Disease Based on Novel Handwriting Markers.” *Ieee Transactions On Neural Systems And Rehabilitation Engineering* 23.3 (2015): 508-16. Web.
- [20] Drotar, Peter, Jiri Mekyska, Irena Rektorova, Lucia Masarova, Zdenek Smekal, and Marcos Faundez-Zanuy.“A New Modality for Quantitative Evaluation of Parkinson’s Disease: In-air Movement.” *13th IEEE International Conference on BioInformatics and BioEngineering* (2013): 1-4. Web.
- [21] Drotár, Peter, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdeněk Smékal, and Marcos Faundez-Zanuy.“Evaluation of Handwriting Kinematics and Pressure for Differential Diagnosis of Parkinson’s Disease.” *Artificial Intelligence In Medicine* 67.C (2016): 39-46. Web.
- [22] Drotar, Peter, Jiri Mekyska, Zdenek Smekal, Irena Rektorova, Lucia Masarova, and Marcos Faundez-Zanuy.“Contribution of Different Handwriting Modalities to Differential Diagnosis of Parkinson’s Disease.” *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings* (2015): 344-48. Web.
- [23] Della Sala, Sergio, Marcella Laiacona, Cristina Trivelli, and Hans Spinnler.“Poppelreuter-Ghent’s Overlapping Figures Test: Its Sensitivity to Age, and Its Clinical Use.” *Archives of Clinical Neuropsychology* 10.6 (1995): 511-34. Web.
- [24] Evelien Nackaerts, Elke Heremans, Bouwien C M Smits-Engelsman, Sanne Broeder, Wim Vandenberghe, Bruno Bergmans, and Alice Nieuwboer.“Validity and Reliability of a New Tool to Evaluate Handwriting Difficulties in Parkinson’s Disease.” *PLoS ONE* 12.3 (2017): E0173157. Web.
- [25] Ejnar Alex Korner, Lise Lauritzen, Flemming Mørkeberg Nilsson, Annette Lolk, and Peder Christensen. “Simple scoring of the clock-drawing test for dementia screening.” *Dan Med J*, 59(1):A4365, 2012.
- [26] San Luciano, Marta, Cuiling Wang, Roberto A. Ortega, Qiping Yu, Sarah Boschung, Jeannie Soto-Valencia, Susan B. Bressman, Richard B. Lipton, Seth Pullman, Rachel Saunders-Pullman, and Weidong Le.“Digitized Spiral Drawing: A Possible Biomarker for Early Parkinson’s Disease.” *PLoS ONE* 11.10 (2016): E0162799. Web.
- [27] Poonam Zham, Dinesh K. Kumar, Peter Dabnichki, Sridhar Poosapadi Arjunan, and Sanjay Raghav.“Distinguishing Different Stages of Parkinson’s Disease Using Composite Index of Speed and Pen-Pressure of Sketching a Spiral.” *Frontiers in Neurology* 8 (2017): 435. Web.

- [28] Julia Kozhenkina. “Quantitative analysis of the kinematic features for the luria’s alternating series test.” Master thesis, Tallinn University of Technology, 2016.
- [29] A. R. Luriaa. “Higher Cortical Functions in Man.” Springer, 1995.
- [30] Burke, Robert E., Stanley, Fahn, Karen, Marder, Lucien, Côté, and Stephen G., Reich. “Evaluation of the Braak Staging Scheme for Parkinson’s Disease: Introduction to a Panel Presentation.” *Movement Disorders* 25.S1 (2010): S76-77. Web.
- [31] Drotár, Peter, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdeněk Smékal, and Marcos Faundez-Zanuy. “Evaluation of Handwriting Kinematics and Pressure for Differential Diagnosis of Parkinson’s Disease.” *Artificial Intelligence In Medicine* 67.C (2016): 39-46. Web.
- [32] Wikipedia contributors. “Anaconda (Python distribution).” Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 8 Aug. 2019. Web. 13 Aug. 2019.
- [33] Sven Nõmm. “Linear model building”. Department of Software Science, Tallinn University of Technology, 7 Mar. 2019: 6. Web.
- [34] community, The SciPy. “Scipy.stats.ttest 1samp”, 2019.
- [35] Ross A., Willson V.L. (2017) “One-Sample T-Test. In: Basic and Advanced Statistical Tests.” SensePublishers, Rotterdam.
- [36] Seber, George Arthur Frederick, and Alan J. Lee. “Linear Regression Analysis.” 2nd ed. New Jersey: Wiley, 2003. Print. Wiley Ser. in Probability and Statistics.
- [37] community, The statsmodels. “statsmodels.” statsmodels, 2019.
- [38] community, The scikit-learn. “scikit-learn cross validation.” cross validation, 2019.
- [39] Wikipedia contributors. “Cross-validation (statistics).” Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 5 Aug. 2019. Web. 14 Aug. 2019.