#2 Regression & Model Evaluation

Summary of the 1st week:

As you could experience, the previous session we started very strong. And it was important to do so because all we saw of data analysis and exploration is 80% of problems and projects.

To make a little summary, we saw:

- A small introduction to machine learning from data, architecture and decision types
- Hypothesis. The important thing to ask yourself before you start exploring the dataset and how to define the problem we have.
- Types of data, missing data, in what ranges our data moves and how to treat temporary data.
- We saw the utility of using histograms, scatterplots, correlations, summary statistics, etc.
- And as you know, we start with the branch of supervised learning. In fact, we will continue with it until the third session.
- We saw the different basic phases of a model: cleaning and data transformation, defining the problem and choosing a model, making a correct tran_test_split, and how to see the score of our model.

Add, that the quality of our model depends in the first instance on the quality of our data. After this, performance concepts come in and how to get performance and adjust the parameters of our model.

- Finally, we focus on a model for classification: decision trees. And we explain that it is the gini index, the entropy, the depth of our tree, etc. (all this is very important for the 3rd session).

We were able to enter more types of algorithms such as kNN, SVM, etc ... However, the idea was to establish the concepts of EDA and the basics of an ML model.

2nd week goals:

This second week we changed from the classification problem to the regression problem. You will see that correlations are more important in these types of problems. And we will begin to see basic concepts of ML:

- Overfitting vs. Underfitting
- Bias vs. Variance
- Regularization
- Scale the dataset
- What is cross-validation?
- How to correctly evaluate an ML model

Contents week 2:

During this week (before Saturday's session), the resources we have chosen are a couple of videos (one more theoretical and one more practical about regression) + a series of readings. **NOTICE**: *We recommend watching the videos, especially the practical one, but if you do not understand very well or do not handle yourself, go directly to the next part.*

The theoretical video is a small introduction to mathematics that we will see in regression. **Our advice is that you spend it more or less quickly depending on your interest**, because then there is much more.

https://youtu.be/ne-MfRfYs_c

And in the practical video, we encourage you to open the attached notebook and follow the steps in the video.

https://youtu.be/B8yIaIEMyIc

And now, after spending more or less time in these two videos. We have attached several notebooks:

- You have a set of notebooks about *linear models and regression* that you should read and follow carefully. **Tip** : go pointing the concepts and doubts.

And as you see, there are many notebooks but short. Once you have gone through them, we are going to the last part of this week.

We have mixed many things this week. It is inevitable to go through feature engineering concepts when talking about regression models, however, we will leave this for next week.

We put some readings to end up understanding all these concepts that we have presented this week. If you have won the 100-Page ML Book, you can read it there 😆

- [Bias vs. Variance](#)

- [Underfitting vs. Overfitting](#)

- [Cross-Validation (CV)](#) ([Cross-validation](#))

- [Regularization](#)

- [Evaluation of a model](#)

Finally: Do you want to do your tests? 👉 [About regression](#) & [About validation](#)

----

https://heartbeat.fritz.ai/classification-model-evaluation-90d743883106

https://towardsdatascience.com/cross-validation-430d9a5fee22

Any questions see you by Slack;) On Saturday we will present a challenge where to apply all these concepts.

RESOURCES

2_4_polynomial_regression.ipynb

2_5_linear_regression_regularization.ipynb

loans.csv

lesson_lasso_ridge.ipynb

2_3_linear_regression_build_multivariate_model.ipynb

2_1_linear_regression-build_univariate_model.ipynb

2_2_linear_regression_check_assumptions.ipynb

#2 Practice!

Good afternoon ☕ Serve your coffee and we start with today's practice!

We are going to leave you some code snippets and tutorials to apply this week's knowledge.

+ [Tutorial](#)

[Diabetes Solution](#)

RESOURCES

[Regression_Analysis.ipynb](#)

#2 Challenge!

In this challenge you will try to predict the Quality variable of the data set.

Once again, we hope you think about what the process you have used returns and draw conclusions.

RESOURCES

[[Solution]wine-prediction.ipynb](#)

[winequality-white.csv](#)

[wine-prediction-Plantilla.ipynb](#)