

Parameters of linear model

Quistian Navarro, Juan Luis
A341807@alumnos.uaslp.mx

January 29, 2024

Ing. Sistemas Inteligentes, Gen. 2021
Machine Learning, Group 281601

Abstract

In the following document, we explain the steps followed, which we will call ‘experimentation’, in the Google colab environment, and the scikit-learn libraries, about the linear regression model[1].

1 Introduction

A model is a representation of the reality that can describe the world in a simplify form.

In this case, will we analyze the Linear Regression. Linear regression is a statistical modeling technique used to describe a continuous response variable as a function of one or more predictor variables.

2 Content

Activities:

1. Run scikit-learn example using Google Colab

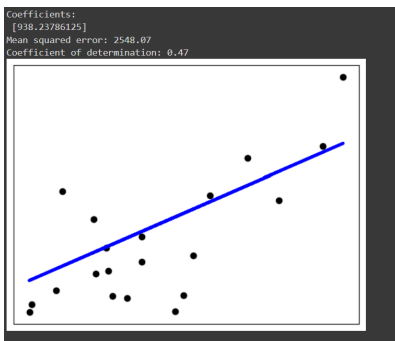


Figure 1: Run scikit-learn example using Google Colab

2. Use 80% for training and 20% for testing and measure performance (error)

In this case, i used this code for split the dataset:

```
# split 80% 20%
diabetes_X_train, diabetes_X_test, diabetes_y_train, diabetes_y_test = train_test_split(
    diabetes_X, diabetes_y, test_size=0.2, random_state=42
)
```

Figure 2: Split 80% training and 20% testing

And this is the output obtain.

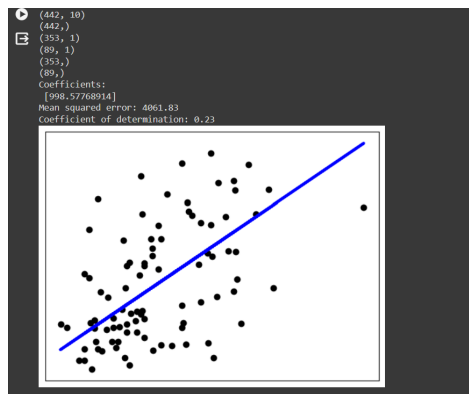


Figure 3: Split 80% training and 20% testing

3. Submit report and your jupyter notebooks

My notebook is here:

https://colab.research.google.com/drive/13Nf4CDgm_uTag5Xt10envQG1xl-OfY2E?usp=sharing

4. Repeat the experiment using accidents.txt dataset (MATLAB Example), see examples below on how to load data in Google Colab

3 Results

For the 'accidents.csv' data set, the notebook is here[2]: <https://colab.research.google.com/drive/1j5xUxIV07ijPVfFJDHFIZyg-CqMEH2Cm?usp=sharing>

And the results obtained are as follows:

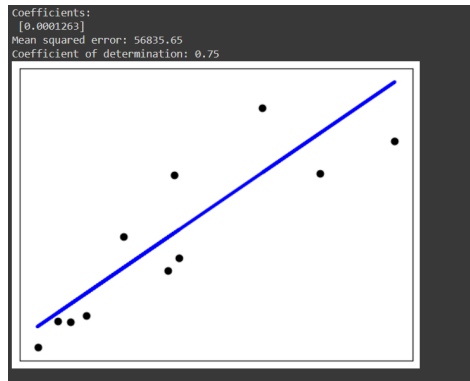


Figure 4: Split 80% training and 20% testing

4 Conclusion

We can mention as a conclusion of the activity, that understanding the use of the libraries offered by Python allows an efficient development of the models, since it is not necessary to focus entirely on how we could implement the entire infrastructure to make a model, on the contrary with using the resources already offered by libraries such as scikit-learn it is possible to do experimentation.

References

- [1] Ordinary least squares.
- [2] Juan Carlos Cuevas-Tello. Handouts on regression algorithms, 09 2020.