MC886 – Machine Learning – T1

Victor Akira Hassuda Silva

106976

Hermes Shimidu

139511

# Introduction

First practice work involving Linear Regression, as objective the model should predict the music year based on a set of variables chosen.

# Activities

Obtain the data set for training and test

Added 1 line on each file to each file to represent the columns and converted the file to csv

Using Python and SciKit + Numpy libraries, we were able to calculate the first linear regression to obtain the 1st coefficients

Apply Cost Function and minimize it to obtain better coefficients

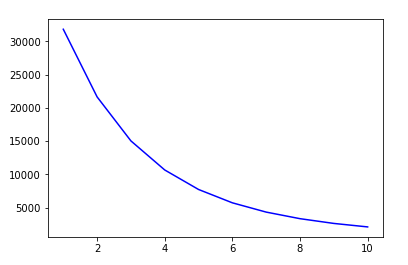
# Proposed Solution

We used multi linear regression to obtain the best prediction and minimized the Cost function to adjust the coefficients for each variable.

# Experiments and Discussion

Before starting, we needed to treat both datasets, first we needed to include one more line into each dataset to represent the column names. After that, for training set, we used Standardization in order to scale all X values and then calculated the Linear Regression using SciKit library. With Linear Regression calculated, we could retrieve the coefficients for each variable analyzed.

Using the first guessing for coefficients, we applied and minimized the Cost Function to obtain better coefficients. For the Cost Function, we used a learning rate of 0.0001 and 10 iterations. Producing the following graph.



We changed the number of variables to analyze, changed the value of learning rate and iteration, but the best model we could achieve wasn’t good enough.

# Conclusion

Variating the number of variables, we obtained different results, taking more variables isn’t the best solution to obtain the best coefficients for linear regression. A better approach is to variate the learning rate and the number of iterations to obtain better coefficients.

# References

"Hands-On Machine Learning with Scikit-Learn and TensorFlow", Aurélien Géron, 2017.