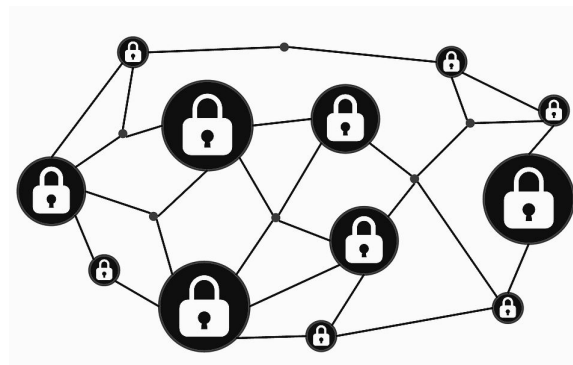universidade
de aveiro

Bitcoin Price Predictor

Foundations of Machine Learning
DETI

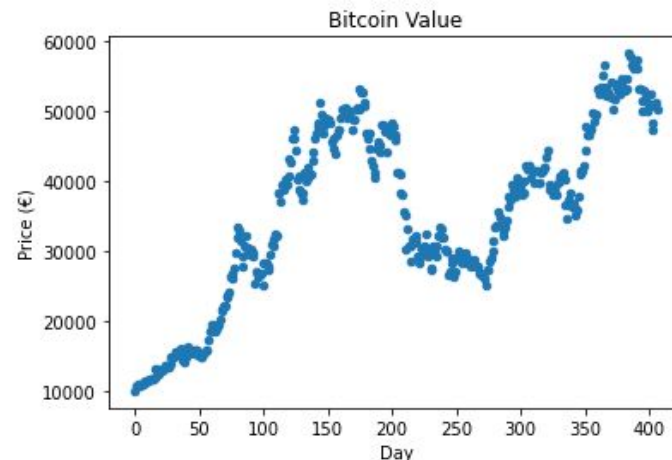Pedro Gonçalves 88859
Rui Oliveira 89216

Introduction

- Motivation
  - Cryptocurrency
  - Blockchain
- State of the art
- Objectives
  - Create an algorithm that predicts the price of Bitcoin for any given day.
  - Find the best model for the minimum error.

Data Description and Pre-Processing

- Dataset
  - Price of Bitcoin of the last 407 days.
- Parsing
  - Divide the data into various parameters/columns, corresponding to the **memory** of the model.
- Normalization
  - Normalize our data to achieve better computation times.

Machine Learning Models

1. Initial Approach
2. Multivariate Regression

$h\theta(x) = \theta_0 + \theta_1 x1 + \theta_2 x2 + \dots$

DAY N-2	DAY N-1	DAY N	DAY Y
10070.59 €	10803.79 €	10985.89 €	10901.23 €
10803.79 €	10985.89 €	10901.23 €	11050.01 €

The last value of each line is the price we try to predict (Y), which will be the price of the last day, in the next line.

3. Difference between prices

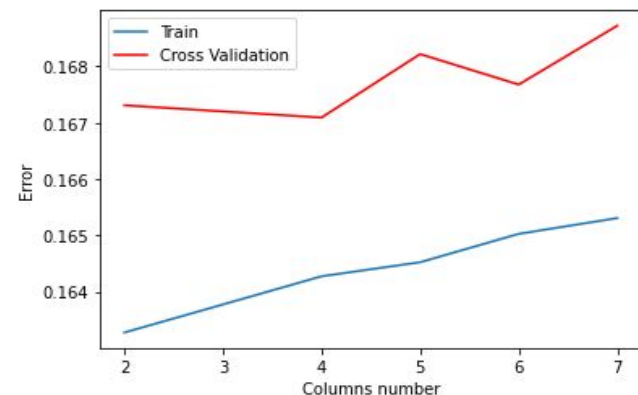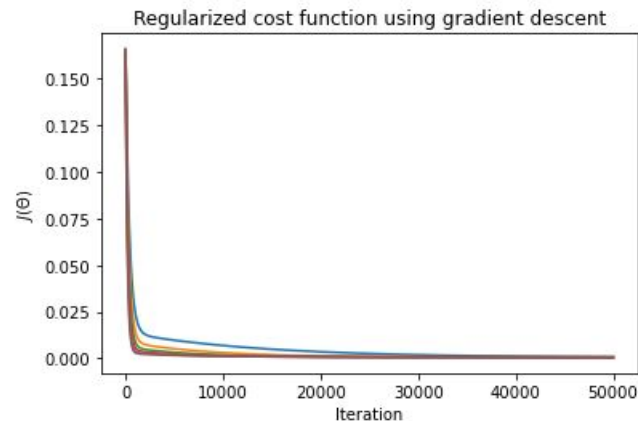| Difference 1	Difference 2	Y
733.2€	182.11€	-84.66€
182.11€	-84.66€	148.78€

The first value of the first line is the difference of the first two days (10803.79 - 10070.59) and so on.
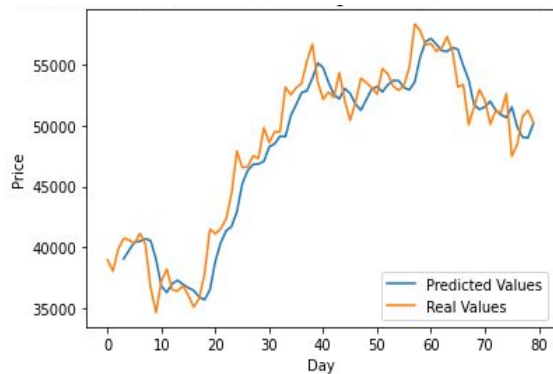
4. Final Model (Mean)

Model training

- Data Splitting
  - 60% - Training data
  - 20% - Validation data
  - 20% - Testing data
- Cost Function
- Ideal parameters
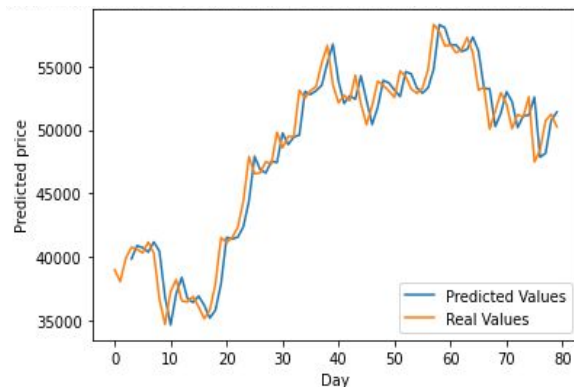 - Find the best number of columns for the model through cross validation.

Regularized cost function using gradient descent

Results

Real Price vs Predicted Price of Testing Data
(Based on last three days)

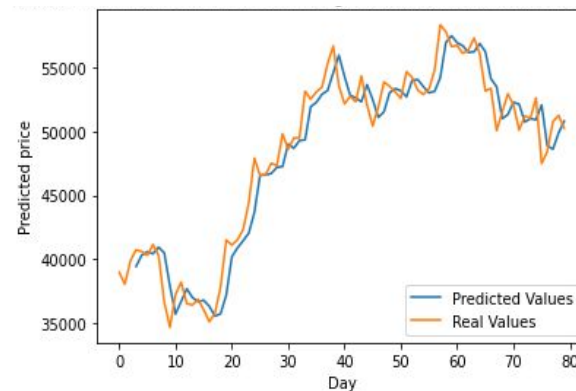h(x) = 0.0112 + 0.2545x + 0.318x2 + 0.4104x3
Price data Error: 0.00052725

Real Price vs Predicted Price of Testing Data
Using differences
(Based on last three days)

h(x) = 0.0021 + 0.055x + -0.0448x2
Difference data Error: 0.00038399

Real Price vs Predicted Price of Testing Data
With the features' mean
(Based on last three days)

Final data Error: 0.00040644

Conclusion

- We concluded that the Bitcoin is highly volatile, which makes it nearly impossible to make an exact guess of the price.

- In the future we intend to increase the number of features, mainly related to news and articles.

- Futurely, we aim to use a different type of dataset based on another time stamp, besides just daily based.