Data Mining

Homework 4

Due: 13/1/2019, 23:59

This homework is not mandatory. You may want to try to do it to practice with the topic and to get extra credit for the course. If you decide to do it, though, you need to hand it it by the due date, to give us time to grade it.

Instructions

You must hand in the homework electronically and before the due date and time.

This homework has to be done by each **person individually**.

Handing in: You must hand in the homework by the due date and time by an email to Leonardo (`martini.1722989@studenti.uniroma1.it`) that will contain as attachment **(not links to some file-uploading server!)** a .zip file with your answers. The filename of the attachment should be `DM_Homework_1__StudentID_StudentName_StudentLastname.zip`;
for example:
`DM_Homework_1__1235711_Robert_Anthony_De_Niro.zip`.
The email subject should be
`[Data Mining] Homework_1 StudentID StudentName StudentLastname`;
For example:
`[Data Mining] Homework_1 1235711 Robert Anthony De Niro`.
After you submit, you will receive an acknowledgement email that your project has been received and at what date and time. If you have not received an acknowledgement email within 2 days after you submit then contact Leonardo.

The solutions for the theoretical exercises must contain your answers either typed up or hand written clearly and scanned.

For information about collaboration, and about being late check the web page.

Problem 1. This problem aims to give you some exposure to the use of neural networks for regression problems. First of all, download the following **Bitcoin Historical Dataset** in csv format from this website:

`https://www.kaggle.com/mczielinski/bitcoin-historical-data`

The **Bitcoin Historical Dataset** contains Bitcoins exchanges for the time period of Jan 2012 to July 2018. This dataset has different columns: the Open, High, Low and Close columns in the data set signify the Opening, highest, lowest and Closing price of Bitcoin against the USD on that particular day. The volume refers to the volume of Bitcoin transferred in the market on a particular day, indicated currency, weighted Bitcoin price and timestamps in Unix time. To have a better overview of the dataset you can visite the following website:

`https://www.kaggle.com/mczielinski/bitcoin-historical-data/home`

You can choose two different tasks to implement:

- **Easier task:** implement Long Short-Term Memory (LSTM) neural network to do the estimation of the price of Bitcoins. To understand how to implement LSTM in keras look at the documentation on the website:

`https://keras.io/layers/recurrent`

- **Harder task**: Given a budget, build a trading strategy that maximizes your profit, investing those money from the 1st January 2018 to the end. Assuming that at every time stamp you can choose to do nothing or to sell or to buy your portfolio.

  For example, a good heuristics might try to buy Bitcoins when you (i.e., your model) believe the price is going to raise, while you should sell when you believe the opposite it is true, namely it is going to fall.

  Note: You should simulate the period of data from January 2018 – end as data that are unknown to you. Thus, should not use this period as a test set. Instead, you use only the data from start – December 2017 as training and test set (you decide where to do the split) and tafter you are happy with that period, you execute the algorithm to the data of January 2018 – end.