ΥΣ19 ARTIFICIAL INTELLIGENCE II (DEEP LEARNING FOR NATURAL LANGUAGE PROCESSING). FALL SEMESTER 2020, HOMEWORK 1

National and Kapodistrian University of Athens, DiT. 20/10/2020

Dimitrios Foteinos | sdi1700181

Contents

1 Batch Gradient Descent, Stochastic Gradient Descent, Mini-Batch Gradient Descent for Linear Regression

In this specific exercise, we were asked to implement:

- Batch Gradient Descent
- Stochastic Gradient Descent
- Mini-Batch Gradient Descent

For this part of homework, we played with California Housing Dataset.

More specifically, I did some Data Preprocessing, Feature Scaling and also I worked with some cool tools and libraries, such as numpy arrays, matplotlib plots, and pandas dataframes. Please, check also my detailed ipynb file for this particular exercise.

***Note**: Our feature engineering didn't worked as we expected. So the scores that we had with these features was not the optimal.

2 Twitter Sentiment Analysis

For this exercise, our job was to use **Ridge Regression**, in order to classify our data, and find either a tweet is **positive** or **negative**.

- Data Proprocessing:

  In this part, we "cleaned up" our data. To be more specific, we transformed all text into lowercase. Then, we removed any **html tags** and all the **special characters.**

- Vectorization method:

  In order to vectorize our tweets into numbers, we use 2 methods: The first was the Bag Of Words, and the second was the Tf-Idf Method.

Finally, we used some measurements to see how much predictable (or accurate) our model is for this project. More specifically, we used:

- Accuracy
- Precision
- Recall
- F-Measure