

Artificial Intelligence - Spring 2019

Assignment #1: Regression **Due Date: 2019, June 5th**

In order to do this assignment, you have to go through regression theories and concepts. In this assignment, you're provided with 2 identical datasets, one normal and one with some outliers. The datasets include information about height, weight and sex (label) of some random people.

Part A: Linear Regression

Datasets: *Data_Normal.txt*, *Data_With_Outlier.txt*

In this assignment, you need to implement **closed form solution**, **stochastic gradient descent (2000 epochs)** and **batch gradient descent (2000 epochs)** algorithms for linear regression on heights (X) and weights (Y). Please perform the following tasks:

1. Normalize the datasets so that the values of each feature change between 0 (for min. value of the feature) and 1 (for max. value of the feature). Note that this task is very important for the desired results of the upcoming tasks.
2. Train each model separately on the normalized datasets and plot the datasets alongside with the obtained regression model. For these plots, the X axis should be the height feature and the Y axis should be the weight feature. In addition, you should discriminate males and females by using different colors.
3. Report your choice of the parameter "*Learning Rate*" for each model and explain the effects of changing this parameter.
4. One of these datasets have some outliers. Does it affect the robustness of the model? Explain.
5. Plot $J(\theta)$ in terms of θ_1 in $[-2:2]$ and θ_2 in $[-2:2]$. (a 3D figure)
6. Explain what does the normalization process do? When would it be useful to normalize the data?

Part B: Weighted Linear Regression

Datasets: *Data_With_Outlier.txt*

The effect of outliers on linear regression method have been analyzed in the previous part. Here, you have to apply weighted linear regression on the dataset which includes outliers.

1. Derive the closed form solution for weighted linear regression(WLR).
2. Propose a weighting function which decreases the effect of outliers with a formula. Explain why it could be appropriate.
3. Normalize the dataset (similar to the previous part).
4. Apply weighted linear regression using your suggested weighting function and find:
 - a. The closed form solution.
 - b. Batch gradient descent solution (2000 epochs).
5. Plot the outlier dataset and the models obtained from both the previous and the current part on the same figure. Compare the results of parts A and B. (2 figures)
6. When and how does WLR work better than simple linear regression?

Notes:

- Your implementation should be functional.
- Prepare your full report in PDF format and include the figures and the answer of the asked questions.
- The allowed programming languages are any language and feel free but my suggestion is Matlab (easy for matrix and vector) or Python
- Assume each row to be a sample in your implementation
- Submit your assignment using a zipped file with the name of “Name_Famlyname_ StdNum.zip” to compuscien@gmail.com with AI-HW#1 subject.