**ADA University, Machine Learning**

**Programming Assignment - 3**

**Linear Regression**

**Dataset.** In this programming homework, we will use a Turbo.az dataset which is a collection of cars (Mercedes C-class). It contains information about 1328 cars with features of (Sheher, Marka, Model, Buraxilish ili, Ban novu, Reng, Muherrikin hecmi, Muherrikin gucu, Yanacaq novu, Yurush, Suretler qutusu, Oturucu, Yeni, Qiymet, Extra Information, Seller’s comment).

Homework is divided into several parts:

**1. Loading data (10%).** Using pandas library in python is recommended.

You will need to read the data from data file (turboaz.csv) and extract only 3 columns for your model:

X1 = “Yurush” (Milage). If samples of “Yurush” are described by string format, remove “km” from the string and convert it to the number format.

X2 = “Buraxilish ili” (Model Year).

Y = “Qiymet” (Price). If prices of car are given in dollar ($) convert them to manat (AZN).

**2. Visualization (10%).** Using matplotlib library (scatter, Axes3D) in python is recommended.

You will need to provide 3 visualizations of data.

1. Qiymet (Y) vs Yurush (X1)
2. Qiymet (Y) vs Buraxilish ili (X2)
3. 3D plot of all three values (Y, X1, X2)

**3. Implementation of Linear Regression from scratch (40%)**

1. Calculate Cost function. Implement a function which returns cost given true y values, x values and coefficients (w).
2. Normalize data using Z score normalization (Recommended).



|  |  |
| --- | --- |
| Mean | Standard deviation |

1. Implement gradient descent algorithm to minimize Cost Function.
   * + Assign initial values of W=(w0, w1, w2) to zero or choose randomly
     + Learning rate: alpha=0.001, you can change it in different experiments
     + Number of iterations: 10000 or you can stop it when two sequential values are too close.
     + Calculate values of parameters using Gradient descent formula.
2. Plot graph of Cost function and describe how it changes over iterations.
3. Plot points of Y (Qiymet) vs X1 (Buraxilish ili) and draw a line of predictions made with parameters you got from gradient descent.

|  |
| --- |
|  |

1. Plot points of Y (Qiymet) vs X2 (Yurush) and draw a line of predictions made with
2. Plot 3D graph of points of Y (Qiymet), X1, X2  and predicted Y (Qiymet) using the same X1 and X2. It should look like this (blue points are true values, red points are predicted values):

|  |
| --- |
|  |

1. **Testing.**

Here are given two new cars which are not in the dataset:

Car 1 { Yurush: 240000, Buraxilish ili: 2000, Qiymet: 11500}

Car 2 { Yurush: 415558, Buraxilish ili: 1996, Qiymet: 8800}

Predict prices of these cars by using your parameters and compare with actual prices.

1. **Linear Regression using library (20%).**

Use a library to fit Linear Regression to the data. You should use the same features (Yurush and Buraxilish ili) as input to this model and perform the same testing (3(h)) as above. (Using scikit-learn library is recommended).

1. **Report (20%).** Write codes, their explanation and achieved results for each step (1-4) of homework in the report. Codes must be in text format. You can take screen shots only for output of program.

**Extra tasks:**

1. Solve linear regression by Normal equation (10 %)
2. Instead of linear function use polynomial which over performed linear hypothesis accuracy (10%)

You are free to use any programming language, although Python is highly recommended. Also, consider using Jupyter Notebook for your own convenience.

**Please follow the instruction below when you submit your programming assignment to the Blackboard System.**

* Submit your homework solution in 2 separate files:

1) Python file (.py or .ipynb)

2) Word file (doc), The Report containing your codes, explanation and achieved results. Write your codes in text format (NOT as images) in the report.

* Write your full name and number of homework assignment in the FILE NAME.
* Do NOT submit files in Zip file format.

You can submit your assignment 3 times before the deadline.

Please be informed that your assignment on coding will be checked through Safe Assign in Blackboard.

Your assignment will be graded based on mentioned criteria if similarity score is less than 50%

If similarity between 50%-75%, your grade will be penalized 50%.

If similarity is more than 75% then your assignment will be reported to Honor Code committee.

**Your assignment will be graded with Zero if any of requirements in the instruction is missing.**

Furthermore, any student whose solution may arise a question, will be asked for some explanations as well.