Ödev

**Assignment I: Linear Regression**
Note: You do not need to return the solutions, however, assignments will be evaluated by asking questions in the exams which will test whether you did the assignment or not.

In this assignment you will implement linear regression and evaluate its performance on the house prices dataset. You can follow the steps below:

1. Write the gradient descent algorithm which takes a data matrix, output values, learning rate, and number of iterations. It should return the parameter values of the linear model.

2. In order to make a simplification, use only the 10 most correlated numeric attributes with SalePrice which are OverallQual, GrLivArea, GarageCars, GarageArea, TotalBsmtSF, X1stFlrSF, FullBath, TotRmsAbvGrd, YearBuilt, and YearRemodAdd.

3. Before running gradient descent normalize the attributes. This is necessary since there are large differences among the magnitudes of the attribute values. For normalization, for every attribute, substract the mean and divide by standard deviation. After this every attribute's mean and std will be 0 and 1, respectively.

4. Run the gradient descent algorithm on the training set. To test convergence, store the cost function's value after every iteration. If you plot these values, you should see a figure like below:
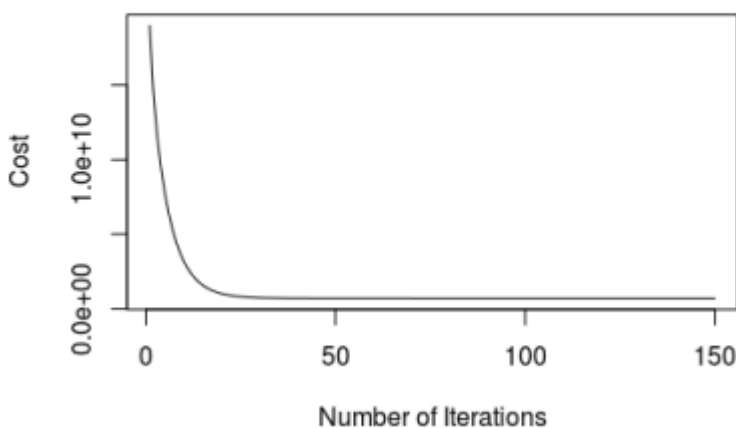
According to this figure, convergence is achieved at about 30th iteration.

5. Split the data randomly into training (80%) and test (20%) sets. Run gradient descent on the training set and evaluate the model on the test set. For evaluation, find Mean Absolute Error (MAE). Run this evaluation several times (for example, 5 times) with different training and test splits, and find average MAE.

6. Once you are a sure that your code is working, you can predict the sale prices of the houses in the test file given in the Kaggle competition, submit your result, and see your ranking.

IMPORTANT NOTE: You are free to use any Python library but you have to implement gradient descent by your self i.e., you cannot use a library function for linear regression.

-Tevfik Aytekin

Bitiş zamanı geçti. Yanıtınızı hala gönderebilirsiniz ancak geç gönderildi olarak işaretlenecek.