Machine Learning HW2 Programming assignment

Gur Asees Singh Chandok[1]

[1]Boston University

February 24, 2018

2.

a. Linear Regression:

This Problem was solved using matlab and the code for the solution is present in file **hw2part2a.m.**

Below derivation was used for getting to the formula for calculating the least squares

In the code $R = I,\ \phi \rightarrow X\ , t \rightarrow y\ and\ w^*\ \rightarrow weights.$ Also,comments are present before each line of code to make it easy to understand.The output of the code is the best choice for 3rd variable.The program would can be used with different data provided the column names are same.However, even if the column names are different the program should give the right index for the best new variable but the label would be incorrect.

Each variable was tested as a prospective 3rd variable and In addition to finding least squares, R2 score was used to choose the variable.

R2

Using this method the variable **'LIC' i.e. 4th column** was chosen as it had the **best R2 score(coefficient of determination) and least squares.**

This can be also be seen in the plot below, the plot shows the R2 score after choosing each variable as 3rd variable on y axis against variable name on the x axis.

Clearly, LIC has the best R2 score after selection as the 3rd variable and therefore it was chosen as the 3rd variable

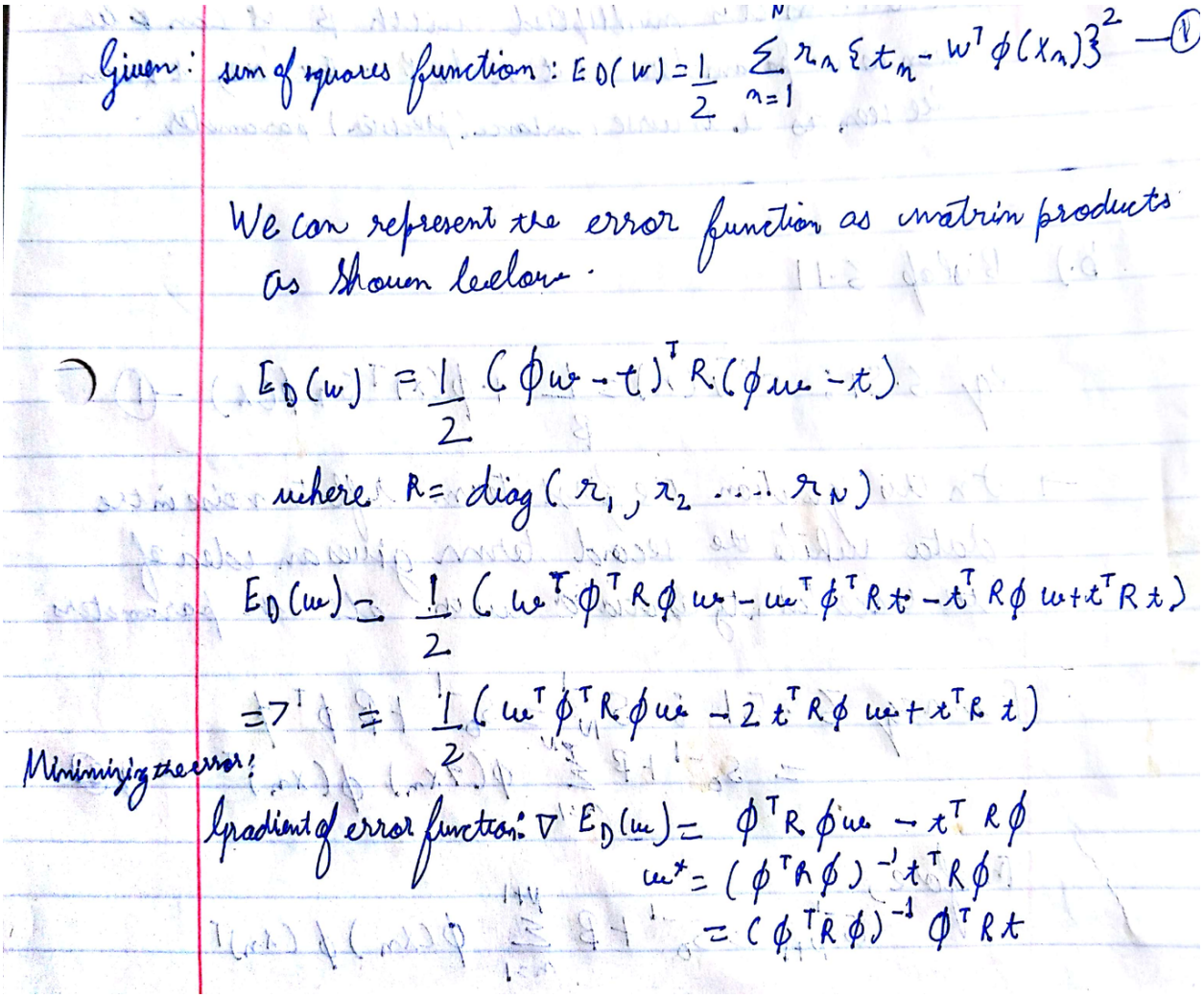

Figure 1: Derivation for computing sum of squares function using Matrix method

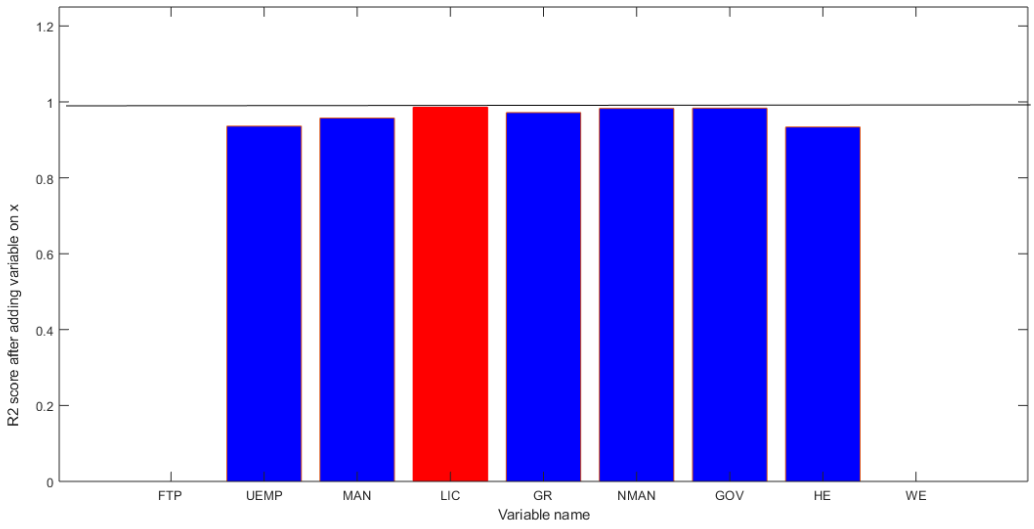

Figure 2: Improvement in R2 score after choosing each variable as variable 3