Johnny Nguyen

September 20, 2022

ECGR 4105: Intro to Machine Learning

Homework #0

GitHub Link: https://github.com/Norumai01/Intro_Machine_Learning/tree/main/HW_0

Problem 1:

- Based on all the linear regression line models that was individually plotted for X1, X2, or X3, X1 seems to have a more accurate regression line for the training model. X2 and X3 had many different training data points that is spread out and the regression line that is plotted would be more ideal.
- X1 had the lowest cost loss as shown as the loss history goes down to 0.985. The lower the cost loss, a better representation of the linear regression line for the training model.
- From testing different ranges of learning on X1, X2, X3, it is found that the best optimal learning rate range around 0.001 – 0.01 depending on the training data points it is given. If the user were the lower the learning rate further, resulted in not identifying the final output of the loss function or requires more iterations to finish it. Increasing the learning rate further just resulted in a strange appearance of the loss over iterations.

Problem 2:

- Variable X1 had a better linear regression model because the regression line is more accurate with the relationship between the X and Y variables.
- A learning rate of 0.01 (if not around 0.008 - 0.01) for the multivariable regression was chosen to have an optimal pacing of the loss over iterations. Any lower would requires more iterations, which is not optimal. Any higher would exponentially decreases and may output inaccurate theta.
- Formula shown at the end of the code in the multivariable regression:
    - H(1,1,1) is 2.9413640816645295
    - H(2,0,4) is 0.09179843075891547
    - H(3,2,1) is -0.012755552822327765