CSE 474/4574: Introduction to Machine Learning
Instructor: Sreyasee DAS BHATTACHARJEE
Homework 1 Due Date - 2/19 12pm
Points - 100

Name - Ryan Xu **Person number -** 50289331

Please provide a short one/two line explanation for multiple choice / true false questions to get full grade.

1. Which of the following methods do we use to find the best fit line for data in Linear Regression? 10pt.

A) Least Square Error
B) Maximum Likelihood
C) Logarithmic Loss
D) Both A and B

Least Square Error since we use a loss function (labled l(w0, w1) on the slides) and minimize the function through optimization (such as through iteration in gradient descent)

2. Suppose that we have N independent variables (X1,X2… Xn) and the dependent variable is Y. Now Imagine that you are applying linear regression by fitting the best fit line using least square error on this data. You found that the correlation coefficient for one of its variables (Say X1) with Y is 0.97.
Which of the following is true for X1? 10 pt.

A) Relation between the X1 and Y is weak
B) Relation between the X1 and Y is strong
C) Relation between the X1 and Y is neutral
D) Correlation can't judge the relationship

The relationship between X1 and Y is strong because it has a correlation coefficient close to 1.

3. True-False: Linear Regression is a supervised machine learning algorithm. 10pt.

A) TRUE
B) FALSE

I think it is supervised because it has defined values for x and y.

4. True- False: Overfitting is more likely when you have huge amount of data to train? 10 pt.

A) TRUE
B) FALSE

False, since with a huge amount of data to train there is less chance to stray off the "intended"/"correct" approximation. ("The larger the dataset, the smaller is the difference between the two.")

5.Which of the following steps in Linear Regression impacts the trade-off between under-fitting and overfitting the most. 10pt.

A. The polynomial degree

This is just mathematics, the higher the degree the more curves it can have (Think the diagram shown in class for overfitting). And when the degree is too small (Think a line on an exponential function) it is underfitting. Compared to the other types this impacts the trade off the most.

B. Whether we learn the weights by matrix inversion or gradient descent

C. The use of a constant-term

6. What is Regression ? 10 pt.

Regression is to give an approximation model for relationships between a set of variables. For example, we used linear regression in order to estimate a line for the set of "Maximum size of the House, affordable for your budget".

7. What is the use of Regularization ? What are L1 and L2 regularizations ? 20 pt.

The use of regularization is to " keep the weights all small, which will prevent any single input dimension to over influence the prediction." This is used to prevent overfitting. L1 regularization "eliminate[s] the weights of the least important features and thus maintains sparsity" by limiting/penalizing through the absolute value of the weights (see lasso regression). L2 regression puts a penalty of the square of weights (see Ridge Regression).

8. Explain Underfitting and Overfitting. Provide and explain ways to prevent overfitting. 20 pt.

Underfitting is when there isn't a good enough approximation for the model (inaccurate). This often happens when there is too little data (think 3 points in a triangle on a graph and trying to put a line through them). Overfitting is when the model tries too hard to find an approximation and does not generalize the data well. One way to prevent overfitting is to have a large set of data, in which case "the larger the dataset, the smaller is the difference between the two" (referring to the correct generalization vs the overfitted generalization). Another way to prevent overfitting is to use prior knowledge/information in order to reduce the set of possible models. Mathematically, regularization can be used to reduce overfitting by limiting/penalizing the influence of any single input dimension (see problem 7 above).