Beryl Sin

PeopleSoft ID: 4522433

02/12/2024

Problem Set 4

1. a) See Reg\_normalEqn.py for funtion’s implementation

b) See ps4.py for code, size of feature matrix (m samples, n features)



c) See ps4.py for code, random seed value was used:A graph of a graph with blue dots and white text

Description automatically generated

A value of 0.001 is suggested for lambda based on these results.

2. a) See ps4.py for code

A graph with blue lines and points

Description automatically generated

A value of 9 is suggested for K based on these results since it corresponds to the highest accuracy. This value is not robust to other problems because it is dependent on the number of samples, complexity, and the spread. Outliers and imbalances in a data set can affect the K value as well.

3. a) See logReg\_multi.py for function’s implementation

b) See ps4.py for code, table:  
 A screen shot of a black background

Description automatically generated

There is some notable difference between the training accuracy and testing accuracy, indicating that there may have been some overfitting (did well on training data, not as well on testing). This could happen as a result of outliers in the data sets.