CS419M Assignment 1 Report

Sai Gaurav Anugole 170070008

February 7, 2020

NOTE: Blocks of code made for each question and correspondingly evaluated by commenting the rest parts

1 Gradient Descent

- **Feature Space:** 4 features were chosen from the given development data namely year, hour, distance(euclidean distance calculated between the coordinates) and number of passengers. This seemed to be the minimum number of features with least RMS obtained over many iterations.
- **RMS obtained:** 5.169559464405932
- **Learning Rate:** 0.25, checked for values $> 0.5 and < 0.1$ where the results seem to diverge. This value reported least rms error (along with stochastic gradient descent) with minimum run time.
- **RMS obtained for closed form solution:** 5.169295161187374
- **Difference between the above two RMS values:** 0.00026430321855830385. This error seems to be very less and hence, Gradient Descent Algorithm also provides features as close to the actual ones as possible.
- **Stopping Criteria** When the absolute value of the gradient of Loss function becomes less than $\epsilon = 10^{-5}$, the corresponding weights obtained then were returned.
- **Difference with stochastic GD:** The RMS error obtained = 11.828491497259924 and the difference with respect to gradient descent = 6.750854988255156. Hence, faster run time but trade off with the RMS error.

2 P-norm

- $\lambda = 10^{-5}$: Higher values reported more RMS error, hence, this value of lambda was chosen
- **RMS error with 2-norm:** 5.169559633959422
- **RMS error with 4 norm:** 5.169559940669643

3 Another Basis functions and L2 regularised model

- **Basis 1:** Previously all the parameters used were linear and normalised between 0 and 1 by simply scaling according to the range of values. The differences with respect to the previous one are:
  1. Square root of hour parameter taken to maximise weights for larger values
  2. Similarly done for the year parameter
  3. Square of number of passengers taken to decrease its effective weight value
- **RMS error using Basis 1:** 5.18112274412604, slightly higher than that compared to the initial choice of basis.

- **Basis 2:** The differences with respect to the initial ones are:
  1. Night hours given more weight conditionally (since data given was in 24h format)
  2. Number of passengers more than 3 only given more weight (believing they don't fit in normal cab, SUV cab with higher rates might have been used)
  3. Instead of taking Euclidean distance, the sum of absolute differences between respective coordinates are taken for distance parameter
- **RMS error using Basis 2:** 5.181124211543557, slightly higher than that compared to the initial choice of basis.

4 Subsets of Training Data

Since Gradient Descent had the least RMS value, it was used for training these 5 subsets and the corresponding RMS errors calculated and plotted.

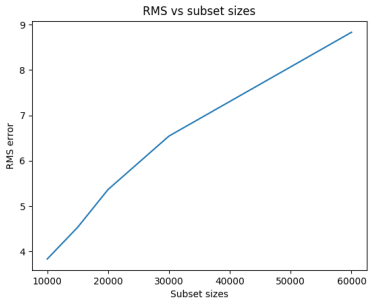

Figure 1: PLOT

5 Least useful features

The Stochastic Gradient Descent seems to be the least useful as it gave large RMS error compared to the rest others with were almost same. The next least useful seemed to be that with the largest subset of training data just because it gave the maximum RMS compared to rest four (although other values evaluated were also very close to it) and gave a large RMS error on the test data

6 Best Regression Model Tested

All choices designed above were tested and Gradient Descent with the initial choice of feature basis on the training data for size of 15000, gave the minimum RMS error of all (5.75)and hence been chosen as the best among all models tested. The differences with the other choices were very small (0.01) and hence also performed well.