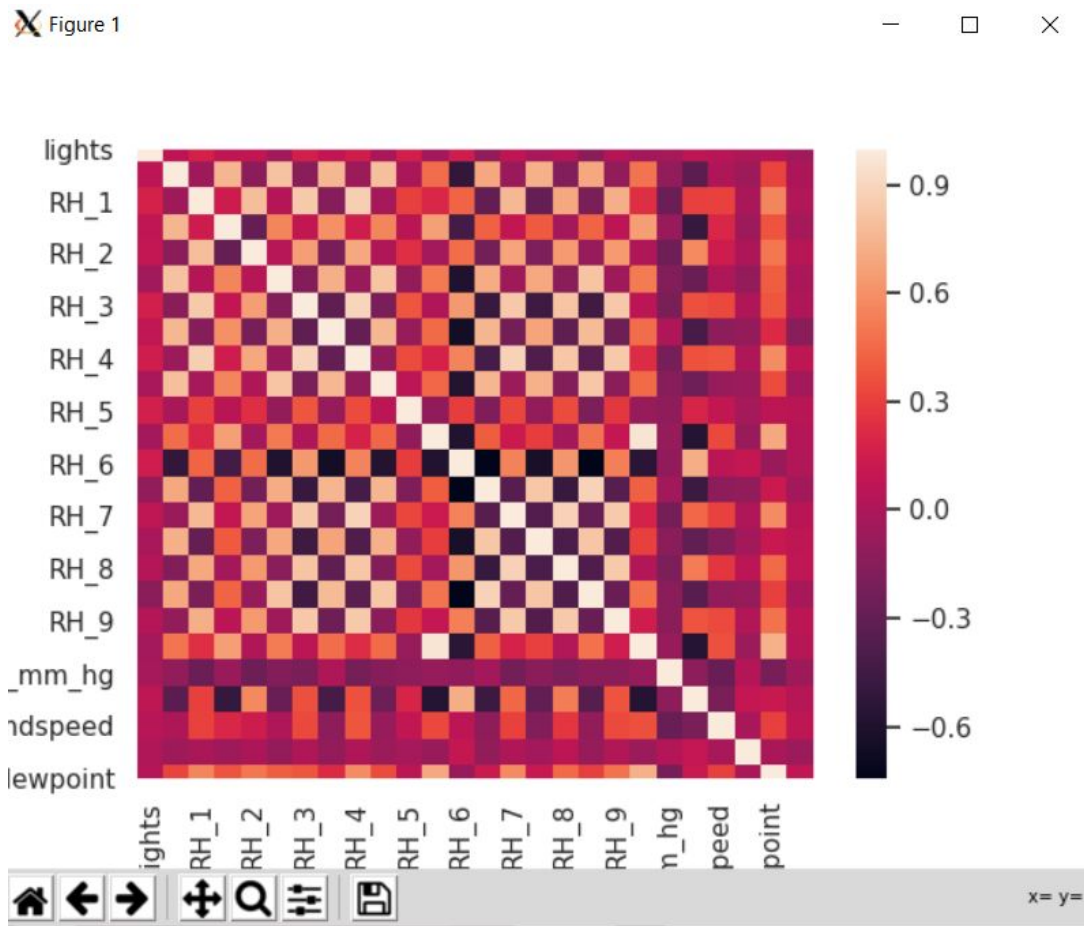


1A: See Code: I extracted the week type from the time column. Weektype means whether the day is during the week or on the weekend. I thought this was useful for energy prediction since on the weekends a person uses more energy since they are at home more often compared to during the week.

1B: See code



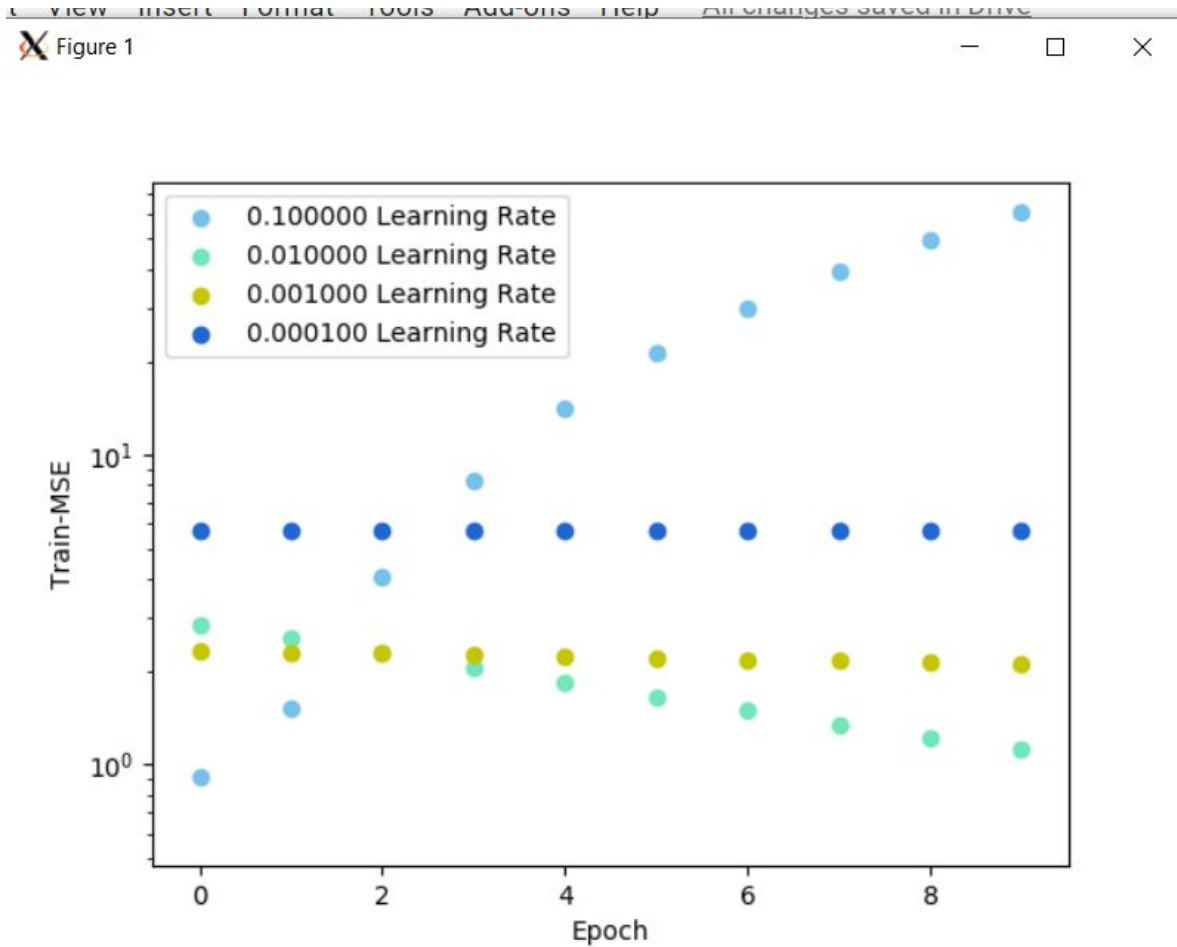
1C: SeeCode: The logic I used for feature selection was to iterate through the correlation matrix and check for the features that hit my target correlation of .75 or greater. I choose .75 as the target correlation because a decent amount of features had correlations above that threshold. I removed one of the features from the pair that had a high correlation until all features left have a correlation less than .75

1D: See Code: For preprocessing the data, I decided to scale the data using min-max scaling. The different features have values and therefore can't be compared to the other features unless the features are scaled

2A-2B: See Code

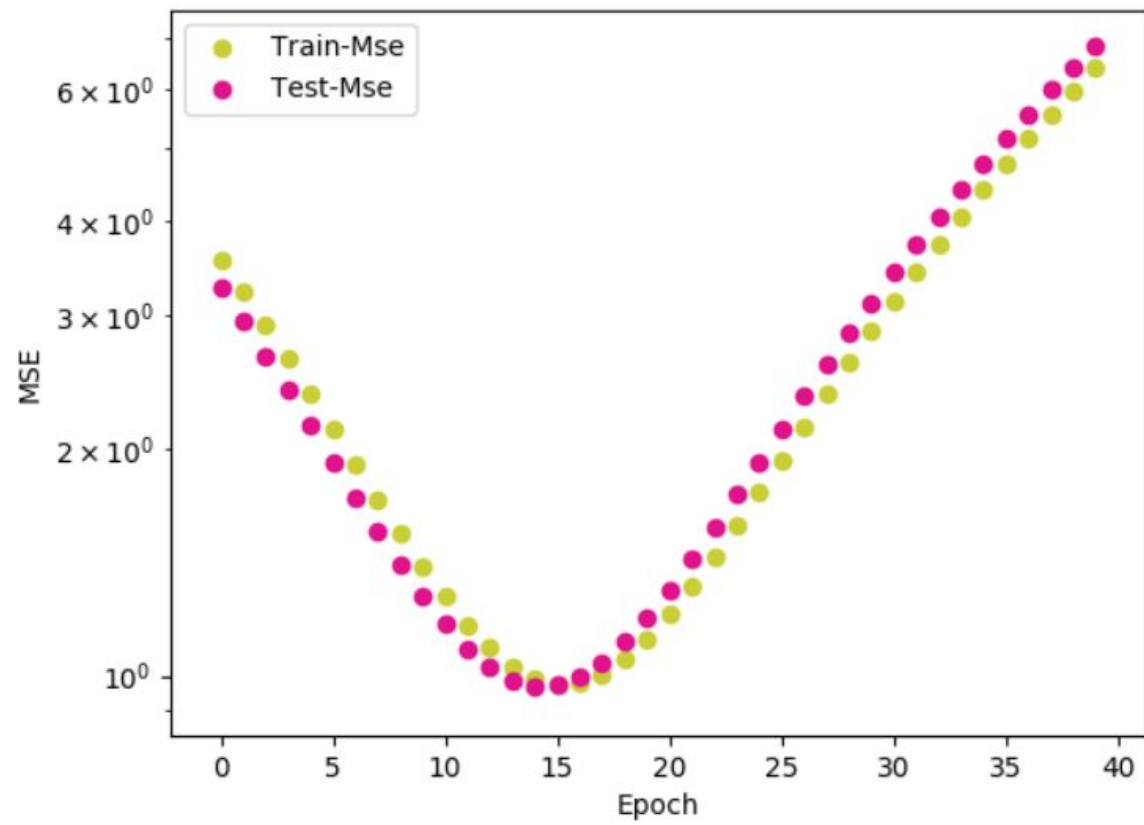
3A: See Code

3B: See Code (find_opt_learn()): I determine the optimal learning rate to be 0.01 for batch size of 1, I looked at the plot and saw where the lowest convergence was.



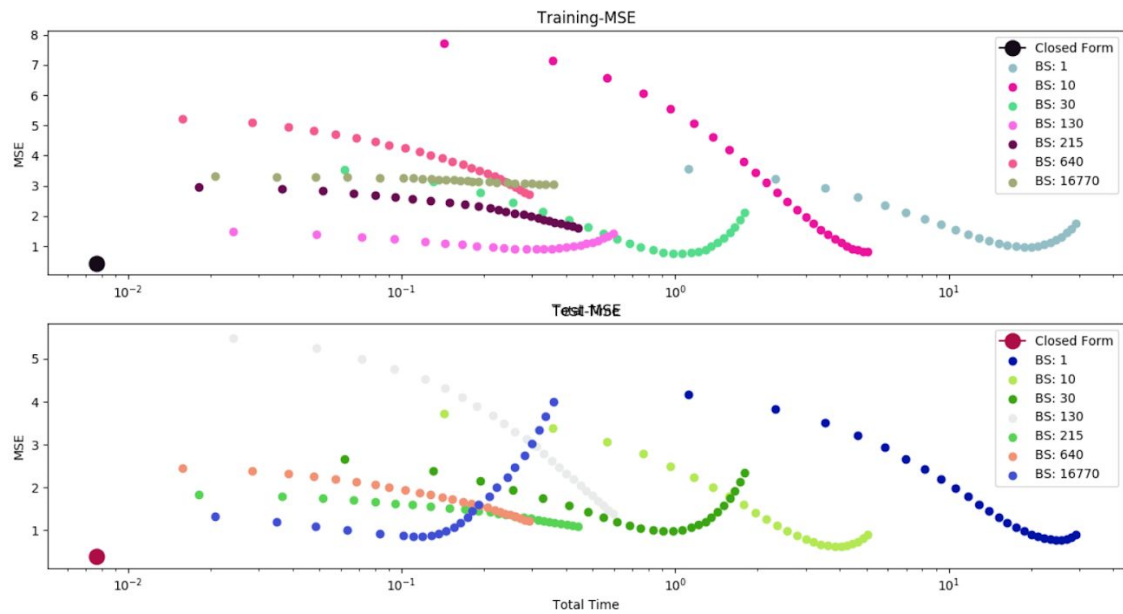
3C: See code (graph_opt_learn()):

Figure 1



4A: See code: comp.py

Figure 1



4B: I noticed that the trade-off between batch size and time. The higher the batch size the data converges in faster time but the trade-off is that the mean square error is higher. The lower the batch size the mse is lower but the time it takes is higher.