CS771A - Assignment 3

1 Question 1

Find out how well can you predict the O3 and NO2 using the method suggested by the manufacturer.Describe the method that gave you the best-performing linear model (in terms of MAE on training data) and write down what mean absolute error (MAE) your model give on the training set.

Answer:
We tried finding our mean absolute error using the different linear models in this part of the question. We used the following three models with their given loss functions and got the MAE for O2 and NO2 as follows:

Ridge Regression using mean square loss with L2 penalty:
Train set:
O2: 5.625980134288402
NO2: 6.540100088345101
Test set:
O2: 6.476017848587366
NO2: 5.754484254431945

Lasso Regression using the L1 loss function:
Train set:
O2: 5.680444323222143
NO2: 6.708687339185759
Test set:
O2: 6.7424969498940595 NO2:
6.16933781903498

Support vector regression which uses -insensitive loss function:
Train set:
O2: 5.648662638237362
NO2: 6.210744036169932
Test set:
O2: 6.037920365632227
NO2: 5.972719046122819

The ridge regression and the support vector regression with the given loss functions produce the lowest MAE.

Preprint. Under review.
Support Vector Regression (SVR) is a variant of Support Vector Machines (SVM) used for regression analysis. SVR works by identifying a hyperplane that maximizes the margin between the predicted values and the actual values. The hyperplane is a decision boundary that separates the predicted values from the actual values. SVR aims to minimize the margin violations, i.e., the difference between the actual and predicted values, while maintaining a tradeoff between the margin size and the number of margin violations.

Ridge Regression is also a linear regression technique that adds a penalty term to the cost function to prevent overfitting. Ridge Regression works by adding a L2 regularization term to the cost function, which reduces the magnitude of the coefficients by forcing them to be closer to zero. The regularization term controls the amount of shrinkage applied to the coefficients.

2 Question 2

Develop a learning method that is free to use temp, humidity, time stamp in addition to the voltage features to predict the O3 and NO2 values useing non-linear models e.g. decision trees, kernels, nearest-neighbors, deep-nets, etc. Describe the method you found to work best giving all details of training strategy e.g. choice of loss function and tuning of hyperparameters.

Answer:
We tried various non linear models like decision trees, kernels, nearest-neighbors, deep-nets, and random forests for this question. We got the lowest MAE and a reasonable time using the support vector regression.

For Support Vector Regression (SVR), the kernel specifies the shape of the decision boundary. Here, the 'rbf' (Radial Basis Function) kernel is used which allows for nonlinear decision boundaries.

Parameter C is the penalty parameter that controls the trade-off between the width of the margin and the errors made by the model on the training data. A smaller C will result in a wider margin, which may lead to underfitting, while a larger C will result in a narrower margin and potentially overfitting to the training data. Here, a value of 100 is chosen, which indicates that the model is willing to tolerate a relatively high number of errors on the training data in order to achieve a narrower margin.

The gamma parameter controls the influence of individual training samples. A small gamma means that each instance has a larger influence on the decision boundary, while a larger gamma means that only nearby instances have an effect. In other words, gamma defines how far the influence of a single training example reaches. Here, a value of 0.1 is chosen, which means that each instance has a moderate influence on the decision boundary.

We use the mentioned parameters to obtain the following time and MAE values:

Train time: 27.74816230580036 MAE
for O2: 0.09930496225817118
MAE for NO2: 0.1171555241546157