CS771 : Assignment 3

Team Members

Anaavi Alok
Anurag Bohra
Dhwanit Balwani
Kartikeya Raghuvanshi
Prakhar Pratap Mall
Ujwal Kumar

1 Linear model to predict O_3 and NO_2 values using just the voltage features

We have used a sklearn to learn a linear model (RidgeCV)
The logic used behind the code can be broken down into the following fundamental steps:

- First, we will seperate the data of ('o3op1', 'o3op2', 'no2op1', 'no2op2') := X to predict 'OZONE':=y1 and 'NO2':=y2.
- Split the data with a test size of 0.2 using train_test_split from sklearn.model_selection.
- Preprocessed the train and test dataset using StandardScaler algorithm
- Next, we use various linear models to train over the training dataset and use MAE as the factor for deciding best model.
- We also perform GridSearchCV to find out the best hyperparameters for the model.
- The table shown below shows values obtained of MAE over the test set

Comparing the MAE values of different models:

Model	Mean Abs Error values, O_3 (MAE)	Mean Abs Error values, NO_2 (MAE)
LinearRegression	5.61	6.59
Ridge	5.61	6.59
LinearSVR	5.59	6.19
Lasso	5.68	6.67
BayesianRidge	5.61	6.59

Table 1: Comparing different linear models

From the values obtained above, we find LinearSVR to be the best model as it gives the minimum values of MAE. The parameters of LinearSVR are: C:10.0 and ϵ=1.0 The parameters of LinearSVR are:

2 Model to predict O_3 and NO_2 with additional features

Here we do not take into account the Time feature into our model. We take in all other features in action. Similar Steps of splitting the dataset and preprocessing is performed and various models are trained. The table below depicts the values of MAE over the test dataset. GridSearchCV is performed

Preprint. Under review.

to find out the best hyperparameters.

Model	Mean Abs Error values (MAE), (combined)
RandomForestRegressor	3.32
XGBRegressor	3.43
LightGBM	3.92
DNN	5.56

Table 2: Comparing different models

As we can observe, RandomForestRegressor gives the best results with MAE of 3.32. The optimal parameters of RandomForestRegressor are n_estimators=200, max_depth=20, min_samples_leaf=1