# CS771 Assignment 3

**Aniruddh Pramod (210142)** **Harsh Bihany (210406)** **Monil Lodha (210630)**

**Akshat Rajani (210812)** **Shreyash Kumar (211009)**

## Abstract

This document describes the methodology and approach used by our group in the third assignment of the course CS771: An Introduction to Machine Learning, offered at IITK in the semester 2022-23-II.

## QUESTION 1

We needed to implement linear models that used the 4 voltage values to obtain the true value of both pollutants, and choose the best-performing linear model in terms of **Mean Absolute Error(MAE)** on training data. We have made the comparison between Linear Regression, Ridge regression and Lasso Regression to come to our conclusion. The following MAE values were obtained :

For OZONE:

| MODEL | MAE |
|---|---|
| ‘ Linear Regression | 6.4735 |
| Ridge Regression | 6.4736 |
| Lasso Regression | 6.5106 |

We see the **"Linear Regression"** model works best for predicting the level of ozone, which implements the **squared error** loss function. We note that the regularization term of Ridge Regression didn’t have significant presence due the small magnitude of our model ’w’.

For NO2:

| MODEL | MAE |
|---|---|
| ‘ Linear Regression | 5.7545 |
| Ridge Regression | 5.7545 |
| Lasso Regression | 5.7293 |

We see the **"Lasso Regression"** model works best for predicting the level of NO2, which implements the **squared error** loss function with L1 regularization.

Preprint. Under review.

## QUESTION 2

We are not using Time as a parameter in our non linear models as the variations in time like diurnal cycle of OZONE is difficult to accomodate leading to increase in the errors of the model.
To accomodate the effects of Temp and Humidity,we tried many different non linear models like KNR(K-Nearest Regression),Kernel-Ridge,Random Forest,Decision Trees using Scikit-learn library and compared the MAE(Mean Absolute Error) and RMSE(Root Mean Squared Error) in each of these models and the results were :
For OZONE:

| MODEL | MAE | RMSE |
|---|---|---|
| KNR,n=1 | 0.0297 | 0.4674 |
| KNR,n=2 | 2.3930 | 3.8998 |
| Kernel Ridge | 6.4736 | 7.6762 |
| Random Forest | 4.0354 | 5.8488 |

We see that the best model we found for predicting OZONE was **K Nearest Regression** with its hyperparameter n (no of neighbors) tuned best at **n=1**.
For NO2:

| MODEL | MAE | RMSE |
|---|---|---|
| KNR,n=1 | 0.0213 | 0.3152 |
| KNR,n=2 | 1.6190 | 3.0974 |
| Kernel Ridge | 5.1712 | 7.2231 |
| Random Forest | 2.9691 | 4.1454 |

We see that the best model we found for predicting NO2 was **K Nearest Regression** with its hyperparameter n (no of neighbors) tuned best at **n=1** After predicting for both NO2 and OZONE we found that **K Nearest Regression** with hyperparamter **n=1** is the best model for predicting the levels of OZONE and NO2.