CS771 Assignment 3

Geetika
Dept. of Comp. Science and Engineering
geetika21@iitk.ac.in

Apoorva Gupta
Dept. of Comp. Science and Engineering
apoorvag21@iitk.ac.in

Girik Maskara
Dept. of Comp. Science and Engineering
girikm20@iitk.ac.in

Amrit Kumar
Dept. of Materials Science and Engineering
amritkr@iitk.ac.in

Akshat Goyal
Dept. of Materials Science and Engineering
akshatgo@iitk.ac.in

Sourabh Mundhra
Dept. of Chemical Engineering
msourabh@iitk.ac.in

Abstract

This document is the submission of group Tompers Squad for Assignment 3. We have answered the parts 1, 2 and 3 with all the relevant level of detail (proofs and tables) required.

1 Part 1

Table 1: MAE for the various methods used

Method	MAE (O_2)	MAE (NO_2)	MAE mean
Least square loss	5.6259	6.5401	6.1830
SGD regressor	5.6327	6.5591	6.0959
Ridge Regression	5.6259	6.5400	6.0830
Lasso Regression	6.4620	7.398	6.9303
Elastic Net Regression	5.6931	6.614	6.1539

Our model performs the best in ridge regression model. Ridge regression model is implemented using scikit learn library in Python.

The model is built using the make_pipeline function from scikit-learn, which sequentially applies data preprocessing steps and model fitting. The data preprocessing step used here is StandardScaler, which standardizes the features by removing the mean and scaling to unit variance.

The hyperparameter alpha of the Ridge regression is set to 0.01. This parameter controls the strength of the regularization, with higher values leading to stronger regularization and smaller coefficients

For the best model, MAE on O_2 = 5.6259 and MAE on NO_2 = 6.5400.

2 Part 2

We used 'no2op1', 'no2op2', 'o3op1', 'o3op2', 'temp' and 'humidity' as the features.

Table 2: MAE and training time for the various methods used

Method	MAE (O_2)	MAE (NO_2)	Training Time(in seconds)
Decision Tree	8.930207	5.06806	0.759
Kernel	0.444978	0.32497	1232.233
Nearest Neighbors	2.392961	1.61779	1.231
Deep-Nets	5.671040	4.98078	3.864

We found KNN(K-Nearest Neighbors) to be the best in terms of MAE and Training time. KNN algorithm predicts the value of a target variable based on the values of one or more input variables. To make a prediction, the algorithm finds the k data points in the training set that is closest to the test data point in terms of distance, where k is a user-defined parameter. The algorithm takes the average value of the target variable among the k neighbors and assigns it as the predicted value for the test data point.

Table 3: Tuning of hyperparameters (varying nearest neighbors) with MAE

Nearest Neighbors	MAE (O_2)	MAE (NO_2)
2	2.39296	1.61779
4	3.00977	2.08837
6	3.28195	2.29216

MAE is least when Nearest Neighbors are two.

Table 4: Loss functions

Loss functions	MAE (O_2)	MAE (NO_2)
MAE	2.392961	1.61779
RMSE	3.901389	3.09726

MAE is the best loss function.

3 Part 3

Code Submitted.