CSE5ML ASSIGNMENT 2

In submitting your work, you are consenting that it may be copied and transmitted by the university for the detection of plagiarism.

“This is my own work. I have not copied any of it from anyone else.”

Name: Maninderpreet Singh Puri

Student number: 20494381

**Python Code**

**KNN regression model:**

from pandas.plotting import scatter\_matrix

import pandas as pd

from matplotlib import pyplot

from sklearn.model\_selection import train\_test\_split

from sklearn.neighbors import KNeighborsRegressor

from sklearn import preprocessing

#Reading the data

filename = 'housing.csv'

names = ['CRIM', 'ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE', 'DIS', 'RAD', 'TAX', 'PTRATIO', 'B', 'LSTAT', 'MEDV']

dataset = pd.read\_csv(filename, delim\_whitespace=True, names=names)

#Describe and Displaying the data

print('First five rows in dataset:', dataset.head())

print('Shape of dataset :', dataset.shape)

print('Type of dataset :', type(dataset))

print('List of columns in dataset :', list(dataset))

print('Length of dataset :', len(dataset))

print('Related Correlational paramteres of dataset :', dataset.describe())

print('Histogram of dataset (column wise):', dataset.hist())

print('Scattered plot:', scatter\_matrix(dataset))

print('Line Graph of dataset: ',dataset.plot())

pyplot.show()

#Train Test Split

from sklearn.model\_selection import train\_test\_split

array = dataset.values

X = array[:,0:13]

Y = array[:,13]

Xx= preprocessing.scale(X)

Yy= preprocessing.scale(Y)

validation\_size = 0.20

seed = 7

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(Xx, Yy, test\_size=validation\_size, random\_state=seed)

#Compile and Fit model

model = KNeighborsRegressor(n\_neighbors=5, metric='l2')

model.fit(X\_train,Y\_train )

print(model)

#Metric

from sklearn.metrics import mean\_squared\_error

predictions = model.predict(X\_test)

print(mean\_squared\_error(Y\_test, predictions))

**SVR regression model:**

from pandas.plotting import scatter\_matrix

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.model\_selection import train\_test\_split

from sklearn.svm import SVR

#Reading the data

filename = 'housing.csv'

names = ['CRIM', 'ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE', 'DIS', 'RAD', 'TAX', 'PTRATIO', 'B', 'LSTAT', 'MEDV']

dataset = pd.read\_csv(filename, delim\_whitespace=True, names=names)

#Describe and Displaying the data

print('First five rows in dataset:', dataset.head())

print('Shape of dataset :', dataset.shape)

print('Type of dataset :', type(dataset))

print('List of columns in dataset :', list(dataset))

print('Length of dataset :', len(dataset))

print('Related Correlational paramteres of dataset :', dataset.describe())

print('Histogram of dataset (column wise):', dataset.hist())

print('Scattered plot:', scatter\_matrix(dataset))

print('Line Graph of dataset: ',dataset.plot())

pyplot.show()

#Train Test Split

from sklearn.model\_selection import train\_test\_split

array = dataset.values

X = array[:,0:13]

Y = array[:,13]

Xx= preprocessing.scale(X)

Yy= preprocessing.scale(Y)

validation\_size = 0.20

seed = 7

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(Xx, Yy, test\_size=validation\_size, random\_state=seed)

#Compile and Fit data

model = SVR(kernel='rbf', gamma='scale', C=44.0, epsilon=0.2, verbose=0)

model.fit(X\_train,Y\_train )

print(model)

#Metric

from sklearn.metrics import mean\_squared\_error

predictions = model.predict(X\_test)

print(mean\_squared\_error(Y\_test, predictions))

**Linear regression model:**

from pandas.plotting import scatter\_matrix

import pandas as pd

from matplotlib import pyplot

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn import preprocessing

#Reading the data

filename = 'housing.csv'

names = ['CRIM', 'ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE', 'DIS', 'RAD', 'TAX', 'PTRATIO', 'B', 'LSTAT', 'MEDV']

dataset = pd.read\_csv(filename, delim\_whitespace=True, names=names)

#Describe and Displaying the data

print('First five rows in dataset:', dataset.head())

print('Shape of dataset :', dataset.shape)

print('Type of dataset :', type(dataset))

print('List of columns in dataset :', list(dataset))

print('Length of dataset :', len(dataset))

print('Related Correlational paramteres of dataset :', dataset.describe())

print('Histogram of dataset (column wise):', dataset.hist())

print('Scattered plot:', scatter\_matrix(dataset))

print('Line Graph of dataset: ',dataset.plot())

pyplot.show()

#Train Test Split

from sklearn.model\_selection import train\_test\_split

array = dataset.values

X = array[:,0:13]

Y = array[:,13]

Xx= preprocessing.scale(X)

Yy= preprocessing.scale(Y)

validation\_size = 0.20

seed = 7

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(Xx, Yy, test\_size=validation\_size, random\_state=seed)

#Compile and Fit data

model = LinearRegression(normalize=True)#normalize=False or true

model.fit(X\_train,Y\_train )

print(model)

#Metric

from sklearn.metrics import mean\_squared\_error

predictions = model.predict(X\_test)

print(mean\_squared\_error(Y\_test, predictions))