Assignment 4 - Problem 2

Divyanshi Kamra, Naman Agrawal, Tushar Nandy, Prasann Viswanathan

April 15, 2020

Abstract

Now that we're clear with what linear regression is, this project is to apply it in order to predict certain scores having other scores as input. Once again the equation use remains the same: (Please note that we've skipped unnecessary chat in this project as the previous was sufficiently in depth and you could use the resources cited in the previous for thorough understanding.)

$$y_{fit} = mx + c \tag{1}$$

1 TOEFL Model Training

To avoid clutter the code has been provided as a link in the bibliography taken from Tushar Nandy's GitHub Page. [1]

2 TOEFL performance

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
#import seaborn as seabornInstance
```

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics

dataframe = pd.read_csv("Admission_Predict_LA.csv")

'''dataframe.plot(x='GREScore', y='TOEFLScore', style='o')
plt.show()'''

x = dataframe['GREScore'].values.reshape(-1,1)
y = dataframe['TOEFLScore'].values.reshape(-1,1)

regressor = LinearRegression()
regressor.fit(x, y) #training the algorithm

#To retrieve the intercept:
c = regressor.intercept_
#For retrieving the slope:
m = regressor.coef_

'''gre = input("Enter GRE Score: ")
gre = float(gre)
toefl = gre*regressor.coef_ + regressor.intercept_
print(f"predicted TOEFL score: {toefl}")'''


y_test = np.array([290, 322, 296, 291, 320, 298, 333, 310, 300,
328, 301, 314, 327, 325, 316, 331, 339, 317, 318, 324])

y_pred = y_test*m + c

df = pd.DataFrame({'Input': y_test.flatten(), 'Predicted': y_pred.flatten()})
print(df)
```

We've included the link for this as well right here- [2]

3 Instruction Manual

All you have to do is download the required resources on your devices. We've provided them as links in the bibliography right here. This was a fairly simple project hence I've kept the instructions brief :) If you come across any difficulty contact @Tushar Nandy or @Prasann Viswanathan through our Whatsapps and we shall do our best to reply promptly.

The sources [3] and [4]

References

Practice
$\texttt{https://github.com/tusharnandy/Team}_LA/blob/master/assgn\ 4/task2 - training.py$

Performance
$\texttt{https://github.com/tusharnandy/Team}_LA/blob/master/assgn\ 4/task2 - performance.py$

TOEFL DataSet
$\texttt{https://github.com/tusharnandy/Team}_LA/blob/master/assgn\ 4/Admission_Predict_LA.csv$

LA DataSet
$\texttt{https://github.com/tusharnandy/Team}_LA/blob/master/assgn\ 4/dataset.csv$