

Assingment 04

August 24, 2021

```
[1]: !pip install matplotlib
```

```
Requirement already satisfied: matplotlib in d:\new folder\new folder\lib\site-packages (3.3.4)
Requirement already satisfied: numpy>=1.15 in d:\new folder\new folder\lib\site-packages (from matplotlib) (1.20.1)
Requirement already satisfied: cyclor>=0.10 in d:\new folder\new folder\lib\site-packages (from matplotlib) (0.10.0)
Requirement already satisfied: python-dateutil>=2.1 in d:\new folder\new folder\lib\site-packages (from matplotlib) (2.8.1)
Requirement already satisfied: kiwisolver>=1.0.1 in d:\new folder\new folder\lib\site-packages (from matplotlib) (1.3.1)
Requirement already satisfied: pillow>=6.2.0 in d:\new folder\new folder\lib\site-packages (from matplotlib) (8.2.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.3 in d:\new folder\new folder\lib\site-packages (from matplotlib) (2.4.7)
Requirement already satisfied: six in d:\new folder\new folder\lib\site-packages (from cyclor>=0.10->matplotlib) (1.15.0)
```

```
[2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
[3]: df=pd.read_csv('agora (1).csv')
```

```
[4]: df.head()
```

```
[4]:
```

	Marketing Spend	Administration	Transport	Area	Profit
0	114523.61	136897.80	471784.10	Dhaka	192261.83
1	162597.70	151377.59	443898.53	Ctg	191792.06
2	153441.51	101145.55	407934.54	Rangpur	191050.39
3	144372.41	118671.85	383199.62	Dhaka	182901.99
4	142107.34	91391.77	366168.42	Rangpur	166187.94

```
[5]: df.shape
```

```
[5]: (50, 5)
```

```
[6]: df.isnull().sum()
```

```
[6]: Marketing Spend    0
      Administration    0
      Transport        1
      Area              0
      Profit           0
      dtype: int64
```

```
[7]: fill=df.Transport.mean()
```

```
[8]: fill
```

```
[8]: 215331.73244897963
```

```
[9]: df.Transport=df.Transport.fillna(fill)
```

```
[10]: df.isnull().sum()
```

```
[10]: Marketing Spend    0
      Administration    0
      Transport         0
      Area              0
      Profit           0
      dtype: int64
```

```
[11]: from sklearn.preprocessing import LabelEncoder
```

```
[12]: le = LabelEncoder()
```

```
[13]: df.Area = le.fit_transform(df['Area'])
```

```
[14]: df.head()
```

```
[14]:   Marketing Spend  Administration  Transport  Area  Profit
0      114523.61      136897.80  471784.10    1  192261.83
1      162597.70      151377.59  443898.53    0  191792.06
2      153441.51      101145.55  407934.54    2  191050.39
3      144372.41      118671.85  383199.62    1  182901.99
4      142107.34       91391.77  366168.42    2  166187.94
```

```
[15]: df.columns
```

```
[15]: Index(['Marketing Spend', 'Administration', 'Transport', 'Area', 'Profit'],
      dtype='object')
```

```
[16]: x = df.drop(['Profit'],axis=1)
```

```
[17]: y = df['Profit']
```

```
[18]: x.shape
```

```
[18]: (50, 4)
```

```
[19]: from sklearn.preprocessing import MinMaxScaler
```

```
[20]: mmx = MinMaxScaler()
```

```
[21]: newx =mmx.fit_transform(x)
```

```
[22]: newx
```

```
[22]: array([[0.69261666, 0.65174393, 1.          , 0.5          ],
        [0.98335946, 0.76197173, 0.94089337, 0.          ],
        [0.92798459, 0.37957895, 0.8646636 , 1.          ],
        [0.87313643, 0.51299839, 0.81223513, 0.5          ],
        [0.85943772, 0.30532804, 0.77613557, 1.          ],
        [0.797566   , 0.3694479 , 0.76912588, 0.5          ],
        [0.81412828, 0.73016111, 0.27071031, 0.          ],
        [0.7880179 , 0.71745725, 0.68649342, 1.          ],
        [0.72901786, 0.74173276, 0.66049977, 0.5          ],
        [0.74590551, 0.43692884, 0.64644319, 0.          ],
        [0.61635061, 0.45150637, 0.48573267, 1.          ],
        [0.60884455, 0.30836422, 0.52936195, 0.          ],
        [0.56766982, 0.57883556, 0.52956308, 1.          ],
        [0.55635219, 0.64106561, 0.53555202, 0.          ],
        [0.72539353, 0.8013272 , 0.54370828, 1.          ],
        [1.          , 0.54302973, 0.55486446, 0.5          ],
        [0.47180821, 0.53527036, 0.56031151, 0.          ],
        [0.57246821, 0.71401273, 0.59894835, 0.5          ],
        [0.55488118, 0.47877201, 0.62511553, 1.          ],
        [0.52264964, 0.77823604, 0.45642007, 0.5          ],
        [0.46116861, 0.47642362, 0.63305328, 0.          ],
        [0.47408436, 0.78021012, 0.63532724, 0.5          ],
        [0.4475048 , 0.54429273, 0.64291963, 1.          ],
        [0.40842369, 0.4146383 , 0.64599195, 1.          ],
        [0.46594728, 0.3653876 , 0.29796428, 0.5          ],
        [0.39107967, 0.67195793, 0.29242745, 0.          ],
        [0.45557444, 0.70684477, 0.28413435, 1.          ],
        [0.43609283, 0.58297807, 0.74861321, 0.5          ],
        [0.39946683, 1.          , 0.25042853, 1.          ],
        [0.39676926, 0.77456642, 0.22709197, 0.5          ],
        [0.37493063, 0.48992809, 0.19316302, 1.          ],
        [0.36974101, 0.77205322, 0.18698856, 0.5          ],
        [0.38348453, 0.5932935 , 0.09768292, 0.          ],
```

```
[0.33561668, 0.39413365, 0.45494286, 1.      ],
[0.2807759 , 0.81005496, 0.44680961, 0.      ],
[0.2782839 , 0.25703165, 0.43561799, 0.5      ],
[0.17335288, 0.57682456, 0.42631115, 1.      ],
[0.26652654, 0.      , 0.41762624, 0.      ],
[0.12234465, 0.11163611, 0.39269043, 0.5      ],
[0.23319442, 0.24130912, 0.3709309 , 0.      ],
[0.17390063, 0.51204073, 0.36626005, 0.      ],
[0.16869099, 0.25446874, 0.34861436, 1.      ],
[0.14297577, 0.34185188, 0.31370517, 0.      ],
[0.09377566, 0.57930693, 0.07531871, 0.5      ],
[0.13412668, 0.78807166, 0.06005866, 0.      ],
[0.0060492 , 0.5547241 , 0.0040356 , 0.5      ],
[0.00795565, 0.49125975, 0.62976785, 1.      ],
[0.      , 0.64054682, 0.      , 0.      ],
[0.00327821, 0.00350184, 0.      , 0.5      ],
[0.      , 0.50014806, 0.09574943, 0.      ]])
```

```
[23]: x=pd.DataFrame(newx,columns=x.columns)
```

```
[24]: x.head()
```

```
[24]:
```

	Marketing Spend	Administration	Transport	Area
0	0.692617	0.651744	1.000000	0.5
1	0.983359	0.761972	0.940893	0.0
2	0.927985	0.379579	0.864664	1.0
3	0.873136	0.512998	0.812235	0.5
4	0.859438	0.305328	0.776136	1.0

```
[25]: from sklearn.linear_model import SGDRegressor
```

```
[26]: from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.30)
```

```
[27]: sgd=SGDRegressor()
```

```
[28]: sgd.fit(xtrain,ytrain)
```

```
D:\New folder\New folder\lib\site-
packages\sklearn\linear_model\_stochastic_gradient.py:1220: ConvergenceWarning:
Maximum number of iteration reached before convergence. Consider increasing
max_iter to improve the fit.
  warnings.warn("Maximum number of iteration reached before "
```

```
[28]: SGDRegressor()
```

```
[29]: sgd.coef_
```

```
[29]: array([95118.08903751, 16976.04156668, 53320.03284153, -690.11403782])
```

```
[30]: sgd.intercept_
```

```
[30]: array([36793.19034639])
```

```
[31]: pred=sgd.predict(xtest)
```

```
[32]: pred
```

```
[32]: array([193432.19190105, 120633.28547961, 122500.87551275, 128161.4016735 ,  
          90508.5113796 , 159592.66704041, 122189.30921138, 104428.4965352 ,  
          130341.14865656, 149627.99188808, 128662.68067319, 94693.83537279,  
          84412.5302702 , 82848.71882837, 66131.73111016])
```

```
[33]: from sklearn.metrics import mean_squared_error
```

```
[34]: mean_squared_error(ytest,pred)
```

```
[34]: 53602215.199077606
```

```
[35]: from sklearn.metrics import r2_score
```

```
[36]: score=r2_score(ytest,pred)
```

```
[37]: score
```

```
[37]: 0.9456598521815915
```

```
[ ]:
```