Assignment 1: Linear Regression

Date: 24-07-2024

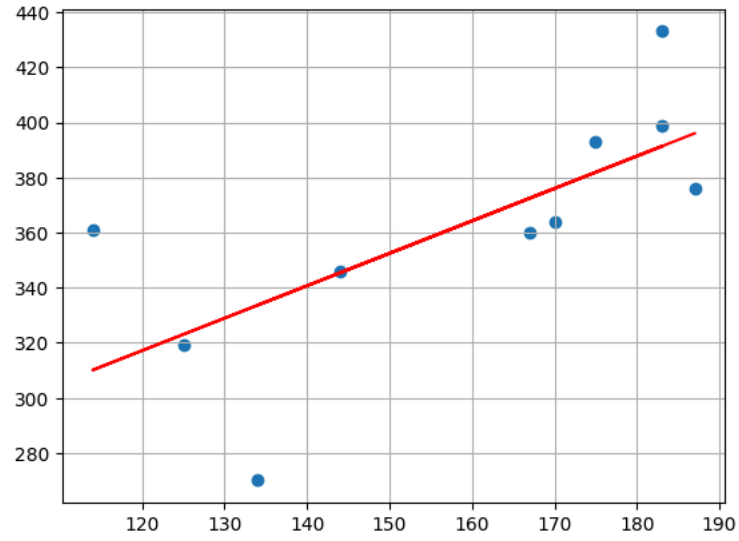
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df=pd.DataFrame(
    np.array([[183,433],[175,393],[134,270],[170,364],[144,346],[183,399],[167,360],[114,361],[125,319],[187,376]]),
    columns=['College Test','University Test'])
print(df)
```

```
   College Test  University Test
0           183              433
1           175              393
2           134              270
3           170              364
4           144              346
5           183              399
6           167              360
7           114              361
8           125              319
9           187              376
```

```
def linear_eqn(df):
    x, y, n = df.iloc[:,0], df.iloc[:,1], df.shape[0]
    m = (n*np.sum(x*y) - np.sum(x)*np.sum(y)) / (n*np.sum(x**2) - np.sum(x)**2)
    b = (np.sum(y) - m*np.sum(x)) / n
    plt.scatter(x, y), plt.plot(x, m*x + b, 'r'), plt.grid(), plt.show()
    return m, b
```

```
m, b= linear_eqn(df)
print("Slope:",m,"\nIntercept:",b)
```




```
Slope: 1.1768620344913772
Intercept: 175.9204261434641
```

```
# Linear Model Prediction
linear_model = lambda x: m * x + b
linear_model(150)
```

```
np.float64(352.4497313171707)
```

```
# MSE Calculation
print("MSE:", np.mean((linear_model(df.iloc[:,0]) - df.iloc[:,1])**2))
```

```
MSE: 927.8656804548871
```

```
# Visualizing the Linear Regression Model
plt.scatter(df.iloc[:,0], df.iloc[:,1])
plt.plot(df.iloc[:,0], m*df.iloc[:,0]+b, 'g')
plt.scatter(150, linear_model(150), c='r', s=100)
plt.xlabel('College Test'), plt.ylabel('University Test')
plt.grid()
```