Shweta_Shinde_Math_Assignment11

December 5, 2024

1 Shweta Ajay Shinde 017548687

MSDA, SJSU , Data 220- Math Method for DA

Find a dataset which has two variables (you will call x and y, but they can have any names). You should have a suspicion or intuition that variable y is linearly dependent on variable x in the dataset.

1. Create a scatter plot of the data (2 pts)

2. Draw the least-square regression line through the scatter plot (3 pts).

3. Calculate and print the coefficient of determination (2 pts) and the correlation coefficient

For this assignment I have select simple dataset with 2 columns Height(Inches) and Weight(Pounds).

```python
[20]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

# Step 1: Load the Dataset
data = pd.read_csv(r'F:/HeightWeight.csv')

x = data['Height'].values.reshape(-1, 1)
y = data['Weight'].values.reshape(-1, 1)

# Step 2: Scatter Plot
plt.scatter(x, y, color='cyan', label='Data Points')
plt.xlabel('Height (Inches)')
plt.ylabel('Weight (Pounds)')
plt.title('Height vs. Weight')

# Step 3:Fit Linear Regression Model
model = LinearRegression()
model.fit(x, y)
y_pred = model.predict(x)
```
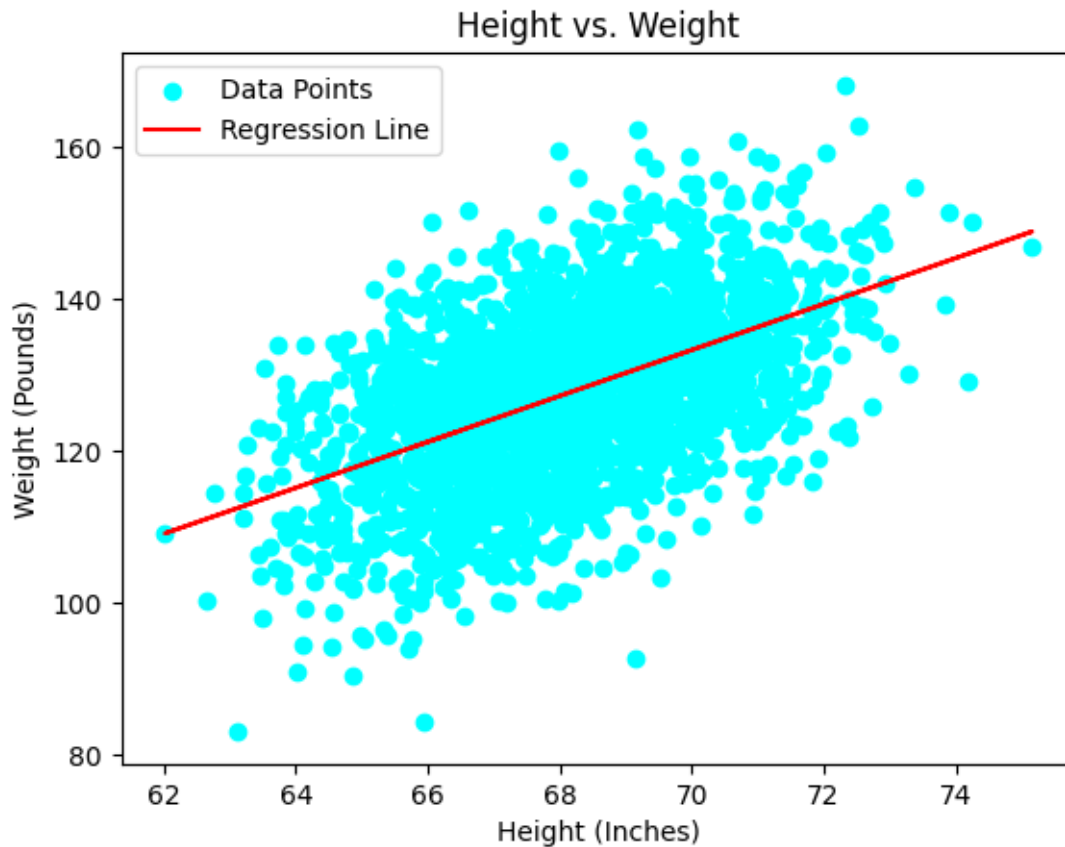
```
# Plot the Regression Line
plt.plot(x, y_pred, color='red', label='Regression Line')
plt.legend()
plt.show()
```


Insights:

In the above scatter plot we see more distribution for Height between 65 to 72 inches and Weight between 120 to 165. The linear Regression line is linearly increasing which suggest if the height increases the weight increases as well.

```
[18]: # Step 4: Calculate R-squared and Correlation Coefficient
      r_squared = r2_score(y, y_pred)
      correlation = np.corrcoef(data['Height'], data['Weight'])[0, 1]

      print(f"Coefficient of Determination (R^2): {r_squared:.2f}")
      print(f"Correlation Coefficient (r): {correlation:.2f}")
```

```
Coefficient of Determination (R^2): 0.26
Correlation Coefficient (r): 0.51
```