Mini Project Report: Height vs Weight Prediction

Introduction

This mini project is about predicting weight based on height using linear regression. We used Python language and scikit-learn library. Random dataset of 20 samples was generated with height in centimeters and weight in kilograms. This project is just for learning purpose not for real medical use.

Objective

Main objective is to learn how linear regression works in machine learning. We tried to find a relation between height and weight so we can predict the weight of a person when height is given.

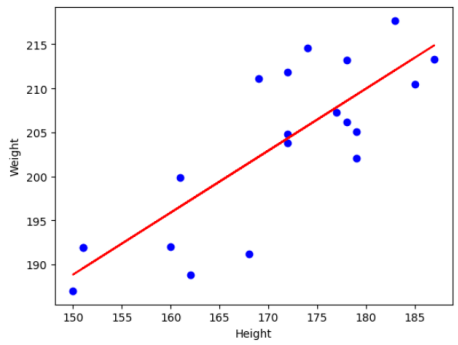
Methodology

1. First generate random data of 20 rows for height (150-190 cm) and weight.
2. Train linear regression model using scikit-learn.
3. Plot scatter plot with regression line.
4. Take user input for height and predict the weight.
5. Compare actual data with predicted data.

Output

Below is the output screenshot of my code execution. It shows the data, regression line, and predicted value for user input height.

```
     h      w
0  178  206.2
1  172  203.8
2  151  191.9
3  172  211.8
4  177  207.3
[0.70411774] 83.22833636513975
```




```
Enter height in cm: 200
Predicted weight: 224.05188523234665
/usr/local/lib/python3.12/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRegression was fitted with featu
  warnings.warn(
```

Observations

From the graph we can see points are not perfect straight but regression line is trying to fit in middle of them. For example, when i entered height 200 cm the model predicted weight around 224 kg which is unrealistic in real life. This shows because we generated random data without real scale. Also model accuracy is not measured here, it is just practice.

Conclusion

In conclusion, we implemented a simple linear regression project for predicting weight using height. Some mistakes are there like weight values not realistic, but we understand the working of regression. In future we can use real dataset and also try multiple regression.

Code

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

np.random.seed(9)
n=20
h=np.random.randint(150,190,n)
w=50+0.9*h+np.random.randint(-10,10,n)
df=pd.DataFrame({"h":h,"w":w})
print(df.head())

X=df[["h"]]
y=df["w"]
reg=LinearRegression().fit(X,y)
print(reg.coef_,reg.intercept_)
pred=reg.predict(X)

plt.scatter(h,w,c="b")
plt.plot(h,pred,"r")
plt.xlabel("Height")
plt.ylabel("Weight")
plt.show()

val=int(input("Enter height in cm: "))
ans=reg.predict([[val]])
print("Predicted weight:",ans[0])
```