01_ML.Simple_linear_regression

June 12, 2023

1 Machine Learning

1.1 Simple Linear Regression

Step 0 Import Libraries

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

Step-1 Import dataset

[ ]:
```
df = pd.read_csv("salary_data.csv")
df.head()
```

[ ]:

	YearsExperience	Salary
0	1.1	39343
1	1.3	46205
2	1.5	37731
3	2.0	43525
4	2.2	39891

Step-2 Spliting dataset into training and testing data

[ ]:
```
X = df[["YearsExperience"]]
y = df["Salary"]
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.2,␣
 ↪random_state=0)
```

Step-3 Fit Linear Regression Model

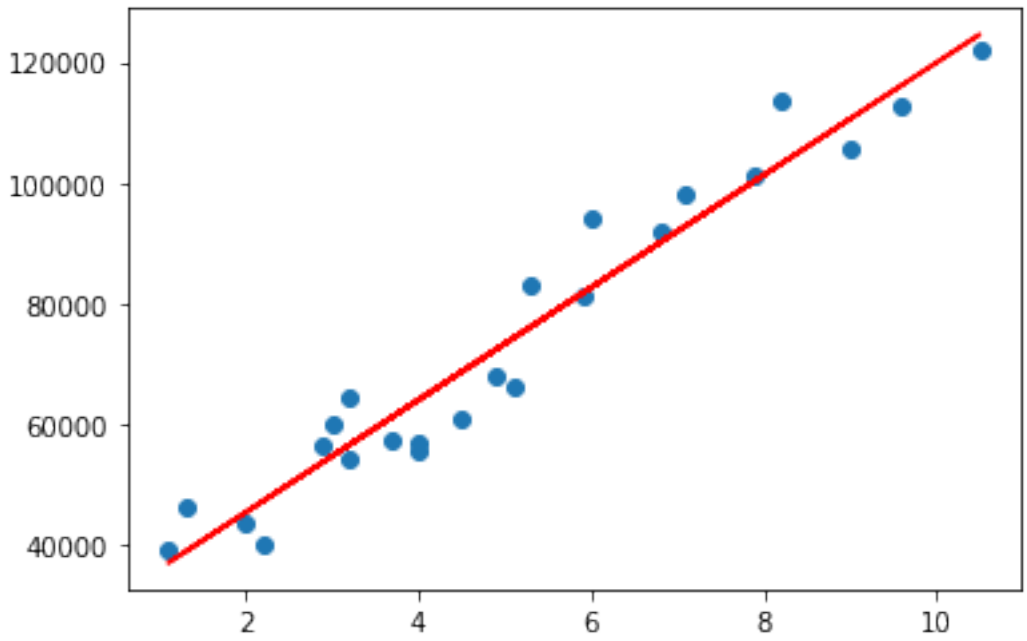
[ ]:
```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model = model.fit(X_train, y_train)
model
```

[ ]:
```
LinearRegression()
```

Step-4 Plotting

[ ]:
```
import matplotlib.pyplot as plt
plt.scatter(X_train,y_train)
plt.plot(X_train.values, model.predict(X_train), color="red")
```

[ ]:
```
[<matplotlib.lines.Line2D at 0x1c28ad80970>]
```



Step-5 Evaluating Model Fitness

The score of the training data is a measure of how well the model performs on the data it was trained on, while the score of the testing data is a measure of how well the model performs on data it has not seen before. The score of the testing data is generally considered to be a more reliable indicator of the model's overall performance, because it reflects the model's ability to generalize to new data.

[ ]:
```
# Model Fitness
print("Score for training data =" ,model.score(X_train, y_train))
print("Score for test data =" ,model.score(X_test, y_test))
```

```
Score for training data = 0.9411949620562126
Score for test data = 0.988169515729126
```

Step-6 Prediction of unknown values

[ ]:
```
model.predict([[10],[15],[20]])
```

```
c:\Users\Saeed Ahmad\AppData\Local\Programs\Python\Python310\lib\site-
packages\sklearn\base.py:450: UserWarning: X does not have valid feature names,
```

```
but LinearRegression was fitted with feature names
  warnings.warn(
```

[ ]:
```
array([119905.85041792, 166468.72605157, 213031.60168521])
```

[ ]:
```

```