simple_linear_regression

January 27, 2022

1 Simple Linear Regression

[ ]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
df = pd.read_csv("salary_data.csv")
df.head(2)
```

[ ]:
```
   YearsExperience  Salary
0              1.1   39343
1              1.3   46205
```

[ ]:
```
X=df[["YearsExperience"]]
y=df['Salary']
```

[ ]:
```
# Import Library
X_train, X_test, y_train, y_test = train_test_split(X,y, test_size=0.2,␣
 ↪random_state=0)
```

1.0.1 Fit linear Regression Model

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

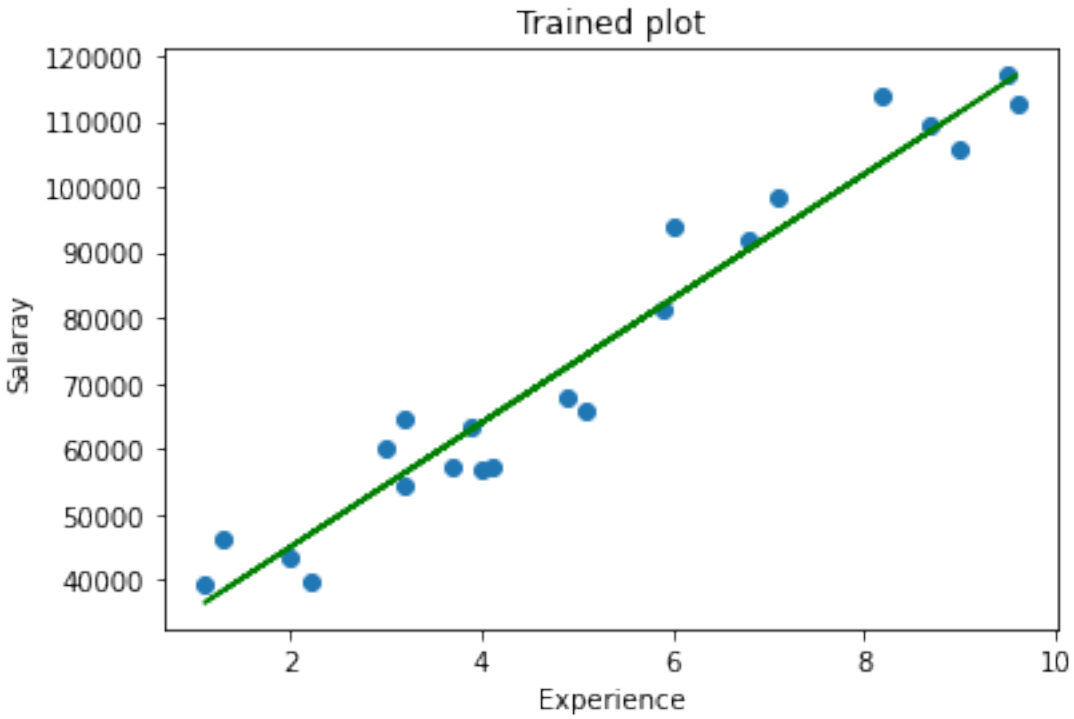
[ ]:
```
LinearRegression()
```

1.0.2 Plotting

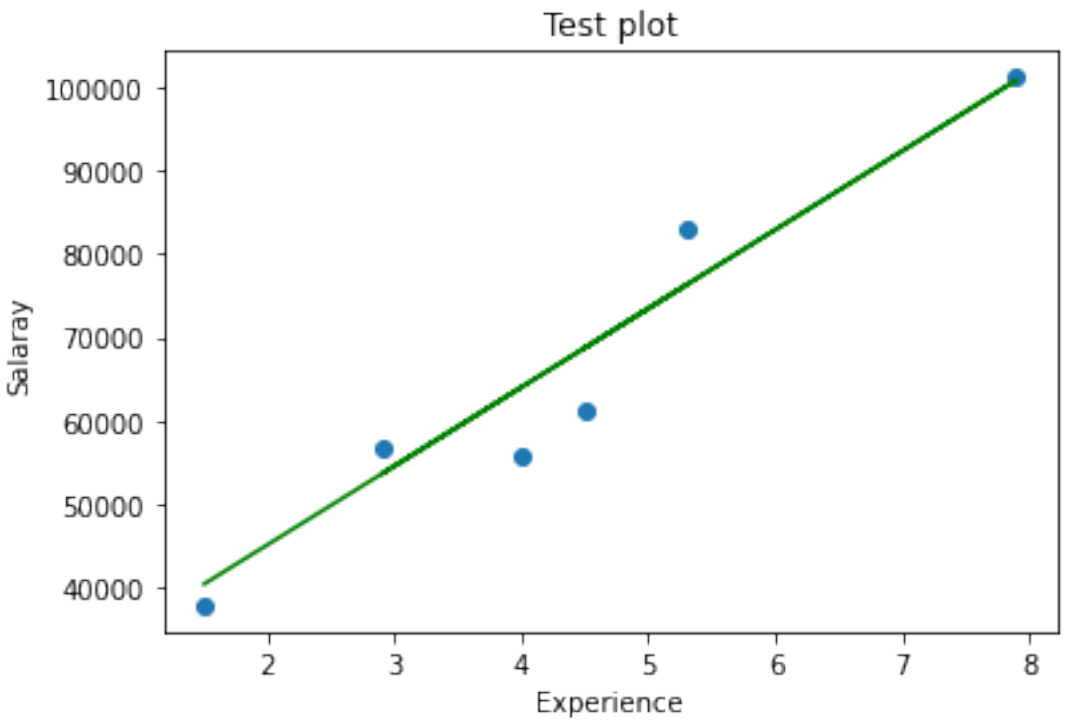
[ ]:
```
import matplotlib.pyplot as plt
plt.scatter(X_train, y_train)
plt.plot(X_train, model.predict(X_train), color='green')
plt.xlabel("Experience")
plt.ylabel("Salaray")
plt.title("Trained plot")
plt.show()
plt.scatter(X_test, y_test)
plt.plot(X_test, model.predict(X_test), color='green')
plt.xlabel("Experience")
```

```
plt.ylabel("Salaray")
plt.title("Test plot")


#plt.scatter(X_test,y_test)
```



[ ]: Text(0.5, 1.0, 'Test plot')



1.0.3 Model Testing/Fitness

[ ]:
```
print("Score for testing data",model.score(X_test,y_test)) # corelation
print("Score for training data",model.score(X_train,y_train))
```

```
Score for testing data 0.9265115445546935
Score for training data 0.9482946812971009
```

1.0.4 Prediction of Unknown Values

[ ]:
```
model.predict([[5]])
```

```
C:\Users\dell7450\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([73476.22072173])
```

[ ]:
```
y_pred = model.predict(X_test)
y_pred
```

[ ]:
```
array([ 40321.21895116, 100947.50790307,  68739.79189737,  76318.07801636,
        53583.21965939,  64003.363073  ])
```