

day2简单线性回归模型

笔记本: 100days

创建时间: 2019/1/10 20:18

URL: <http://localhost:8888/notebooks/Day2/day2%E7%AE%80%E5%8D%95%E7%BA%BF%E6%80%A7%E5%9B...>

Simple Linear Regression(简单线性回归模型)¶

Step 1: Data Preprocessing(数据预处理)¶

In [6]:

```
x
```

```
import pandas as pd
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
dataset = pd.read_csv('studentscores.csv')
```

```
dataset
```

Out[6]:

```
   HOURS  SCORES  
0      2.5     21
```

HOURS	SCORES	
1	5.1	47
2	3.2	27
3	8.5	75
4	3.5	30
5	1.5	20
6	9.2	88
7	5.5	60
8	8.3	81
9	2.7	25
10	7.7	85
11	5.9	62
12	4.5	41
13	3.3	42
14	1.1	17
15	8.9	95
16	2.5	30
17	1.9	24
18	6.1	67
19	7.4	69
20	2.7	30
21	4.8	54
22	3.8	35
23	6.9	76
24	7.8	86

...

In [7]:

```
X = dataset.iloc[:,1].values
```

```
Y = dataset.iloc[:,1].values
```

```
X,Y
```

Out[7]:

```
(array([[2.5],
        [5.1],
        [3.2],
        [8.5],
        [3.5],
        [1.5],
        [9.2],
        [5.5],
        [8.3],
        [2.7],
        [7.7],
        [5.9],
        [4.5],
        [3.3],
        [1.1],
        [8.9],
        [2.5],
        [1.9],
        [6.1],
        [7.4],
        [2.7],
        [4.8],
        [3.8],
        [6.9],
        [7.8]]),
 array([21, 47, 27, 75, 30, 20, 88, 60, 81, 25, 85, 62, 41, 42, 17, 95, 30,
        24, 67, 69, 30, 54, 35, 76, 86], dtype=int64))
```

...

In [9]:

```
x
```

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=1/4, random_state=0)
```

```
X_train, X_test, Y_train, Y_test
```

Out[9]:

```
(array([[7.8],
       [6.9],
       [1.1],
       [5.1],
       [7.7],
       [3.3],
       [8.3],
       [9.2],
       [6.1],
       [3.5],
       [2.7],
       [5.5],
       [2.7],
       [8.5],
       [2.5],
       [4.8],
       [8.9],
       [4.5]]], array([[1.5],
       [3.2],
       [7.4],
       [2.5],
       [5.9],
       [3.8],
       [1.9]]], array([86, 76, 17, 47, 85, 42, 81, 88, 67, 30, 25, 60, 30, 75, 21, 54, 95,
       41], dtype=int64), array([20, 27, 69, 30, 62, 35, 24], dtype=int64))
```

...

Step 2: Fitting Simple Linear Regression Model to the training set(训练集使用简单线性回归模型来训练)¶

In [15]:

```
xxxxxxxxxx
```

```
from sklearn.linear_model import LinearRegression
```

```
regressor = LinearRegression()
```

```
regressor = regressor.fit(X_train, Y_train)
```

```
regressor
```

Out[15]:

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,  
                 normalize=False)
```

...

Step 3: Predecting the Result(预测结果)¶

In [18]:

```
2
```

```
Y_pred = regressor.predict(X_test)
```

```
Y_pred
```

Out[18]:

```
array([16.84472176, 33.74557494, 75.50062397, 26.7864001 , 60.58810646,  
       39.71058194, 20.8213931 ])
```

...

Step 4: Visualization(可视化)¶

Visualising the Training results(训练集结果可视化)¶

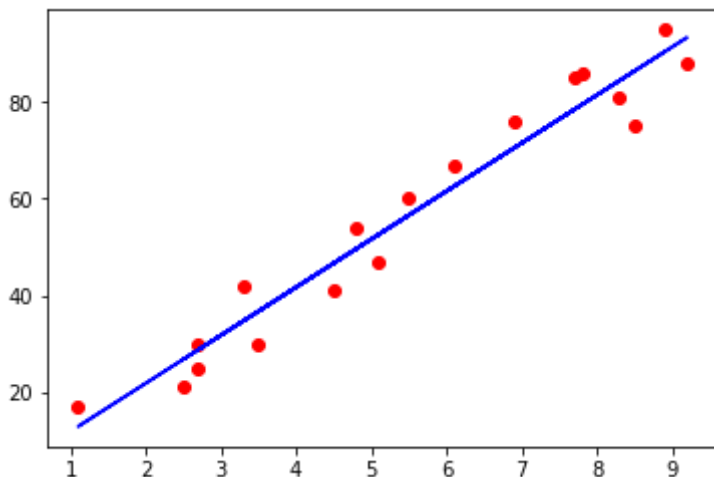
In [19]:

3

```
plt.scatter(X_train, Y_train, color='red')
```

```
plt.plot(X_train, regressor.predict(X_train), color='blue')
```

```
plt.show()
```



...

Visualizing the test results(测试集结果可视化)¶

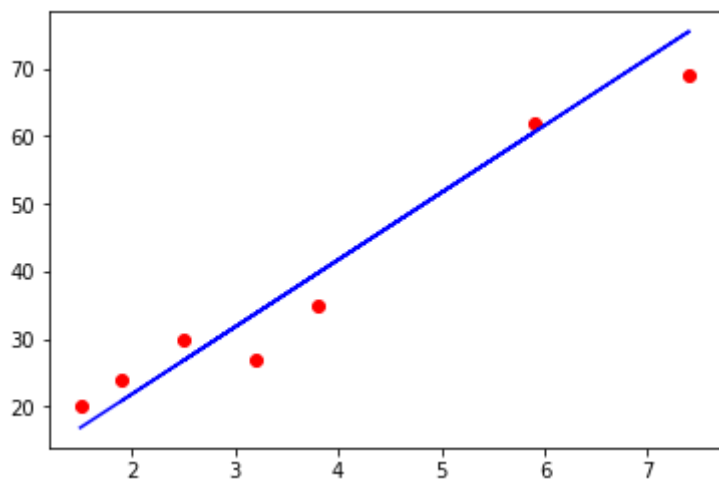
In [22]:

3

```
plt.scatter(X_test, Y_test,color='red')
```

```
plt.plot(X_test, regressor.predict(X_test), color='blue')
```

```
plt.show()
```



...

In []:

1



...