# BÀI TUẦN 02: HỔI QUY TUYẾN TÍNH ĐƠN BIẾN (SIMPLE LINEAR REGRESSION)

### 1. Thông tin sinh viên

Họ tên: Dương Minh Lượng

MSSV: 18521071

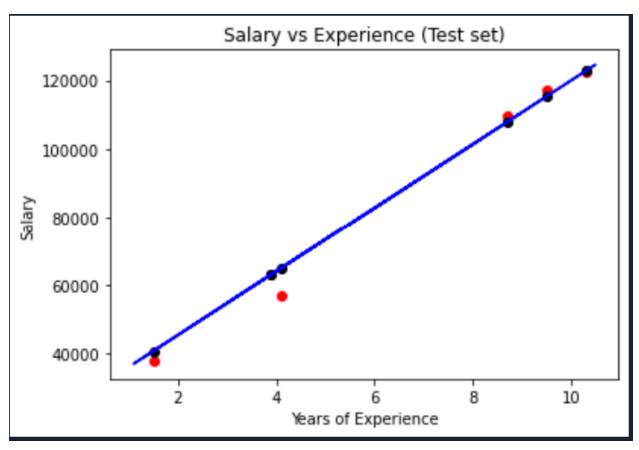
LÓP: Học máy thống kê-DS102.L12.CNCL

#### 2.Source

```
1. import pandas as pd #cho du lieu tu file
2. import numpy as np
                       #xu li mang
3. import matplotlib.pyplot as plt #truc quan hóa dữ liệu
4. from sklearn.model selection import train test split #phân
  chia dữ liêu
5. from sklearn.linear model import LinearRegression
6. from sklearn.metrics import r2 score
7. #Tiền xử lí dữ liệu
8. dataset =pd.read csv("Salary Data.csv")
9. X=np.array(dataset.iloc[:,:-1].values)
       Y=np.array(dataset.iloc[:,1].values)
11.
       X train, X test, Y train, Y test=train test split(X, Y, s
  huffle=True ,train size=0.8, random state=0)
     reg=LinearRegression()
12.
13.
       reg.fit(X train, Y train)
14.
       Y train pred=reg.predict(X train)
15.
      Y test pred=reg.predict(X test)
16.
       #hoàn thiên
      plt.scatter(X test, Y test, color = 'red')
17.
18.
      plt.scatter(X test, Y test pred, color="BLACK")
      plt.plot(X train,Y train pred, color="BLUE")
19.
20.
      plt.title('Salary vs Experience (Test set)')
      plt.xlabel('Years of Experience')
21.
22.
      plt.ylabel('Salary')
23.
      plt.show()
24.
       def compare(i example):
    x=X_test[i_example : i_example+1]
25.
26.
      y=Y test[i example]
27.
       y pred=req.predict(x)
28.
      print(x,y,y pred)
29.
       for i in range(len(X test)):
30.
       compare(i)
31.
      #Đánh giá mô hình
32.
      print('(R)2 train = ',reg.score(X train, Y train))
       print('(R)2 test = ',reg.score(X_test, Y_test))
33.
      r2=r2 score(Y test,Y test pred)
34.
```

```
35.
        print('(R)2 test = ',r2)
36.
        if ((reg.score(X train,
  Y train)>=0.8) & (reg.score(X test, Y test)>=0.8)):
37.
            print('Mô hình tốt')
38.
        elif ((reg.score(X train,
  Y train) == 1) & (reg.score(X test, Y test) == 1)):
39.
             print('Mô hình cơ sở')
40.
        else:
            print('Can xem lai')
41.
```

## 3. Kết quả



```
In [18]: runfile('F:/MAY HQC/THBudi2/BaiThucHanh.py', wdir='F:/MAY HQC/THBudi2')
[[1.5]] 37731.0 [40748.96184072]
[[10.3]] 122391.0 [122699.62295594]
[[4.1]] 57081.0 [64961.65717022]
[[3.9]] 63218.0 [63099.14214487]
[[9.5]] 116969.0 [115249.56285456]
[[8.7]] 109431.0 [107799.50275317]
(R)2 train = 0.9411949620562126
(R)2 test = 0.988169515729126
(R)2 test = 0.988169515729126
Mô hình tốt

In [19]:

IPython console History
```

#### Nhân xét:

- Khi dùng với việc lấy kết quả ngâu nhiên thì R>0.8 -> Mô hình tốt
- Các Y\_test thực tế và Y\_test\_pred dự đoán gần nhau các điểm gần nhau .
- Nếu mà ta dùng với việc shuffle=False Thì mô hình không được tốt vì do dữ liệu xếp năm kinh nghiệm theo lương tăng dần hình minh hoa dưới đây là R

```
In [13]: runfile('F:/MAY HOC/THBuối2/BaiThucHanh.py', wdir='F:/MAY HOC/THBuối2')
[[8.7]] 109431.0 [111119.08832991]
[[9.]] 105582.0 [114134.92418014]
[[9.5]] 116969.0 [119161.31726387]
[[9.6]] 112635.0 [120166.59588062]
[[10.3]] 122391.0 [127203.54619784]
[[10.5]] 121872.0 [129214.10343133]
(R)2 train = 0.9179154343152582
(R)2 test = 0.07028895951395653
(R)2 test = 0.07028895951395653
Cần xem lại

IPython console

History
```