

BÀI TUẦN 03: MULTIPLE LINEAR REGRESSION

1. Thông tin sinh viên

DƯƠNG MINH LƯỢNG-18521071

2. Source

```
1. import time
2. star = time.time()
3. import pandas as pd
4. import numpy as np
5. import itertools
6. from sklearn.linear_model import LinearRegression
7. # Importing the dataset
8. dataset = pd.read_csv('BostonTrain.csv')
9. dataset1 = pd.read_csv('BostonTest.csv')
10. Y_train=np.array(dataset.iloc[:, -1].values)
11. Y_test=dataset1.iloc[:, -1].values
12. K=[0,1,2,3,4,5,6,7,8,9,10,11,12]
13. re=[]
14. fd=[]
15. for i in itertools.combinations(K, 5):
16.     X_test=dataset1.iloc[:,[i[0],i[1],i[2],i[3],i[4]]].values
17.     X_train=dataset.iloc[:,[i[0],i[1],i[2],i[3],i[4]]].values
18.     lin_reg = LinearRegression()
19.     lin_reg.fit(X_train, Y_train)
20.     lin_reg.score(X_train, Y_train)
21.     lin_reg.score(X_test, Y_test)
22.     re.append((lin_reg.score(X_test, Y_test)))
23.     fd.append(i)
24.
25. dfs = pd.DataFrame(fd)
26. list_f=list(dataset.columns)
27. del list_f[-1]
28. dfs=dfs.replace(K,list_f)
29. dfs.insert(5,'R2',re)
30. dfs.to_csv("result10.csv",index=False)
31. print("Tổng thời gian chạy chương trình là ", time.time()-star)
```

3. Kết quả

dfs - DataFrame						
Index	0	1	2	3	4	R2
0	ZN	INDUS	CHAS	PTRATIO	LSTAT	0.402921
1	CHAS	NOX	DIS	PTRATIO	LSTAT	0.399367
2	ZN	CHAS	TAX	PTRATIO	LSTAT	0.394272
3	INDUS	CHAS	TAX	PTRATIO	LSTAT	0.393474
4	INDUS	CHAS	NOX	PTRATIO	LSTAT	0.387926
5	ZN	CHAS	NOX	PTRATIO	LSTAT	0.387467

Cấu hình máy: ASUS ROG STRIX GL502VT

RAM : 16GB

Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz

SSD :120 GB

GPU : Intel(R) HD Graphics 530 và NVIDIA Geforce GTX 970M

Tổng thời gian chạy chương trình là = 2.217073678970337(s)

Nhận xét:

- Với 1287 vòng lặp thì cho được thời gian chạy là 2.2s và với cấu hình trên -> cũng không nhanh lắm.
- Với việc chọn 5 trong 13 biến độc lập thì R cho kết quả thấy chỉ với 0.4 là cao nhất còn có cả R âm -> vì do X_test và Y_train là 2 biến phụ thuộc thì kết quả theo 13 biến độc lập.
- Trong 5 kết quả R(2) cao nhất thì các biến phụ thuộc là LSTAT, PTRATIO, CHAS là xuất hiện trong cả 5 lần->đây là 5 biến độc lập có ảnh hưởng đến Y nhất và cần được quan tâm kĩ.
- Kết quả R thấp nhất là -2.2024 của 5 biến phụ thuộc là CRIM, NOX, RM, RAD, B