介紹:

資料集採用房屋價格的資料，本實驗要根據**建物移轉總面積平方公尺、建物現況格局-房**、**建物現況格局-廳**、**建物現況格局-衛**，id出**房屋總價格**

**程式碼 (python):**

# -\*- coding: utf-8 -\*-

import numpy as np

import matplotlib.pyplot as plt

#from sklearn.preprocessing import StandardScaler

import pandas as pd

from numpy.linalg import inv

from sklearn.model\_selection import train\_test\_split

df = pd.read\_csv('A\_LVR\_LAND\_A.csv', encoding='big5')

A = np.zeros([df.shape[0], df.shape[1]-1])

Y = np.zeros([df.shape[0], 1])

c = 0

for key in df.keys():

if '總價元' == key:

Y[:,0] = df[key]

print('Y : ', key)

else:

A[:,c] = df[key]

c+=1

print('X-', c, ' : ', key)

Y = Y/1e6 # 除以百萬

### Generate data myself ( Testing )

#A1 = np.linspace(1,100,1000).reshape(-1,1)

#A2 = np.linspace(50,100,1000).reshape(-1,1)

#A = np.hstack((A1, A2))

#Y = 3.14\*A[:,0] - 1.65\*A[:,1]

#Y = Y.reshape(-1, 1)

# Function Y = A \* theta

# theta = (A.T \* A)^-1 \* A.T \* Y

# 後面 20% 作測試

A\_train, A\_test, Y\_train, Y\_test = train\_test\_split(A, Y, test\_size=0.2, shuffle=True)

theta = np.matmul((inv(np.matmul(A\_train.T, A\_train))), np.matmul(A\_train.T, Y\_train))

print('theta ', theta)

pred\_Y\_train = np.matmul(A\_train, theta)

error\_train = np.mean(np.abs(Y\_train - pred\_Y\_train))

print('Error train : ', error\_train)

dot\_x = np.linspace(1, Y\_train.shape[0], Y\_train.shape[0])

plt.figure(figsize=(20,12))

plt.plot(dot\_x, Y\_train, 'b^--')

plt.plot(dot\_x, pred\_Y\_train, 'r>--')

plt.xlabel('dot')

plt.ylabel('million($)')

plt.title('training (80%)')

plt.legend(['True', 'Predict'])

plt.show()

pred\_Y\_test = np.matmul(A\_test, theta)

error\_test = np.mean(np.abs(Y\_test - pred\_Y\_test))

print('Error test : ', error\_test)

dot\_x = np.linspace(1, Y\_test.shape[0], Y\_test.shape[0])

plt.figure(figsize=(20,12))

plt.plot(dot\_x, Y\_test, 'b^--')

plt.plot(dot\_x, pred\_Y\_test, 'r>--')

plt.xlabel('dot')

plt.ylabel('million($)')

plt.title('testing (20%)')

plt.legend(['True', 'Predict'])

plt.show()

**輸出結果:**

X- 1 : 建物移轉總面積平方公尺 X- 2 : 建物現況格局-房 X- 3 : 建物現況格局-廳 X- 4 : 建物現況格局-衛 Y : 總價元 theta (參數)

[[ 0.16480567] [-0.20675151] [-0.29511331] [ 1.7830075 ]] Error train : 8.539183583931818 Error test : 13.366800335294633

圖片橫軸為第幾筆數據，縱軸為房屋價格(unit : 百萬)，藍色是實際data，紅色為id過後預測出來的結果。由圖中可以發現，system id出來的結果跟實際的結果趨勢變化接近。

