

Python



基礎資料分析應用

-Simple Linear Regression-

pyt294150 Yishan Cheng

2018.2.9

-STEP 1-

安裝第三方套件

{ **numpy** 科學計算套件
Scipy 科學計算套件

scikit-learn (sklearn) 機器學習套件

matplotlib 資料視覺化套件

Pandas 數據處理、資料分析

安裝第三方套件

<https://pypi.python.org/pypi> → 下載原始檔 → 解壓縮 → pip 安裝...

```
C:\WINDOWS\system32\cmd.exe

C:\Users\user>pip install scikit-learn
Collecting scikit-learn
  Downloading scikit-learn-0.19.1-cp36-cp36m-win32.whl (3.9MB)
    100% |#####| 3.9MB 70kB/s
Installing collected packages: scikit-learn
Successfully installed scikit-learn-0.19.1

C:\Users\user>pip install numpy
Collecting numpy
  Downloading numpy-1.14.0-cp36-none-win32.whl (9.8MB)
    100% |#####| 9.8MB 45kB/s
Installing collected packages: numpy
Successfully installed numpy-1.14.0

C:\Users\user>pip install Matplotlib
Collecting Matplotlib
  Downloading matplotlib-2.1.2-cp36-cp36m-win32.whl (8.5MB)
    100% |#####| 8.5MB 32kB/s
Collecting cycler>=0.10 (from Matplotlib)
  Downloading cycler-0.10.0-py2.py3-none-any.whl
Requirement already satisfied: python-dateutil>=2.1 in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from Matplotlib)
Collecting pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 (from Matplotlib)
  Downloading pyparsing-2.2.0-py2.py3-none-any.whl (56kB)
    100% |#####| 61kB 660kB/s
Collecting pytz (from Matplotlib)
  Downloading pytz-2017.3-py2.py3-none-any.whl (511kB)
    100% |#####| 512kB 433kB/s
Requirement already satisfied: numpy>=1.7.1 in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from Matplotlib)
Requirement already satisfied: six>=1.10 in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from Matplotlib)
Installing collected packages: cycler, pyparsing, pytz, Matplotlib
Successfully installed Matplotlib-2.1.2 cycler-0.10.0 pyparsing-2.2.0 pytz-2017.3

C:\Users\user>pip install pandas
Collecting pandas
  Downloading pandas-0.22.0-cp36-cp36m-win32.whl (8.2MB)
    100% |#####| 8.3MB 52kB/s
Requirement already satisfied: numpy>=1.9.0 in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from pandas)
Requirement already satisfied: python-dateutil>=2 in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from pandas)
Requirement already satisfied: pytz>=2011k in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from pandas)
Requirement already satisfied: six>=1.5 in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from python-dateutil>=2->pandas)
Installing collected packages: pandas
Successfully installed pandas-0.22.0
```

安裝第三方套件

```
C:\WINDOWS\system32\cmd.exe
100% | 512kB 433kB/s
Requirement already satisfied: numpy>=1.7.1 in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from Matplotlib)
Requirement already satisfied: six>=1.10 in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from Matplotlib)
Installing collected packages: cyclor, pyparsing, pytz, Matplotlib
Successfully installed Matplotlib-2.1.2 cyclor-0.10.0 pyparsing-2.2.0 pytz-2017.3

C:\Users\user>pip install pandas
Collecting pandas
  Downloading pandas-0.22.0-cp36-cp36m-win32.whl (8.2MB)
100% | 8.3MB 52kB/s
Requirement already satisfied: numpy>=1.9.0 in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from pandas)
Requirement already satisfied: python-dateutil>=2 in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from pandas)
Requirement already satisfied: pytz>=2011k in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from pandas)
Requirement already satisfied: six>=1.5 in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from python-dateutil>=2->pandas)
Installing collected packages: pandas
Successfully installed pandas-0.22.0

C:\Users\user>pip install SciPy
Collecting SciPy
  Downloading scipy-1.0.0-cp36-none-win32.whl (26.0MB)
100% | 26.0MB 21kB/s
Requirement already satisfied: numpy>=1.8.2 in c:\users\user\appdata\local\programs\python\python36-32\lib\site-packages (from SciPy)
Installing collected packages: SciPy
Successfully installed SciPy-1.0.0

C:\Users\user>
```

-STEP 2-

程式內容

1

```
import numpy as np
import matplotlib.pyplot as plt
temperatures = np.array([29, 28, 34, 31, 25, 29, 32, 31, 24, 33, 25, 31, 26, 30])
ice_pop_sales = np.array([80, 65, 96, 85, 58, 64, 80, 75, 58, 92, 52, 75, 65, 85])

from sklearn.linear_model import LinearRegression #載入線性迴歸模組，訓練並建立線性模型
lm = LinearRegression() #lm: linear model #自變數:氣溫temperatures/依變數:冰棒銷量ice_pop_sales
lm.fit(np.reshape(temperatures, (len(temperatures), 1)),
        np.reshape(ice_pop_sales, (len(ice_pop_sales), 1)))
print('係數=%.3f'%(lm.coef_)) # 印出迴歸係數
print('截距=%.3f'%(lm.intercept_))# 印出截距
```

2

```
predict1 = np.array([30]) #測量實際氣溫，預測30度時冰棒會賣出多少量
psales1 = lm.predict(np.reshape(predict1, (len(predict1), 1)))
predict2 = np.array([15]) #預測氣溫15度時冰棒會賣出多少量
psales2 = lm.predict(np.reshape(predict2, (len(predict2), 1)))
print('氣溫30度時預測的冰棒銷量為%.3f'%(psales1)) # 印出氣溫30度時所預測的冰棒銷量
print('氣溫15度時預測的冰棒銷量為%.3f'%(psales2)) # 印出氣溫15度時所預測的冰棒銷量
```

3

```
#開始繪製散佈圖(scatter plot)
plt.scatter(temperatures, ice_pop_sales, color='blue')
plt.plot(temperatures, lm.predict(np.reshape(temperatures, (len(temperatures), 1))),
         color='black', linewidth=1.25)
plt.plot(predict1, psales1, color = 'red', marker = 's', markersize = 7)
plt.xlabel('temperatures') #xlabel=x軸文字/ylabel=y軸文字
plt.ylabel('ice_pop_sales')
plt.show()
```

分析結果

1-建立線性迴歸分析模型

用Python建立一個簡單線性迴歸模型(Simple Linear Regression Model, lm)
→以氣溫預測冰棒的銷量

```
import numpy as np
import matplotlib.pyplot as plt
temperatures = np.array([29, 28, 34, 31, 25, 29, 32, 31, 24, 33, 25, 31, 26, 30])
ice_pop_sales = np.array([80, 65, 96, 85, 58, 64, 80, 75, 58, 92, 52, 75, 65, 85])

from sklearn.linear_model import LinearRegression #載入線性迴歸模組，訓練並建立線性模型
lm = LinearRegression() #lm: linear model #自變數:氣溫temperatures/依變數:冰棒銷量ice_pop_sales
lm.fit(np.reshape(temperatures, (len(temperatures), 1)),
       np.reshape(ice_pop_sales, (len(ice_pop_sales), 1)))
print('係數=%.3f'%(lm.coef_)) # 印出迴歸係數
print('截距=%.3f'%(lm.intercept_))# 印出截距

===== RESTART: C:/Users/user/Desktop/f_py_lm.py =====
係數=3.900
截距=-40.079
>>> |
```

分析結果

2-以線性迴歸分析模型進行預測

```
predict1 = np.array([30]) #測量實際氣溫，預測30度時冰棒會賣出多少量
psales1 = lm.predict(np.reshape(predict1, (len(predict1), 1)))
predict2 = np.array([15]) #預測氣溫15度時冰棒會賣出多少量
psales2 = lm.predict(np.reshape(predict2, (len(predict2), 1)))
print('氣溫30度時預測的冰棒銷量為%.3f'%(psales1)) # 印出氣溫30度時所預測的冰棒銷量
print('氣溫15度時預測的冰棒銷量為%.3f'%(psales2)) # 印出氣溫15度時所預測的冰棒銷量
```

```
===== RESTART: C:/Users/user/Desktop/f_py_lm.py =====
氣溫30度時預測的冰棒銷量為76.914
氣溫15度時預測的冰棒銷量為18.417
>>>
```

分析結果

3-資料視覺化: 散佈圖

#開始繪製散佈圖(scatter plot)

```
plt.scatter(temperatures, ice_pop_sales, color='blue')  
plt.plot(temperatures, lm.predict(np.reshape(temperatures, (len(temperatures), 1))),  
         color='black', linewidth=1.25)  
plt.plot(predict1, psales1, color='red', marker='s', markersize=7)  
plt.xlabel('temperatures') #xlabel=x軸文字/ylabel=y軸文字  
plt.ylabel('ice_pop_sales')  
plt.show()
```

