



## 第8章

### 機器學習實戰（一）： 用線性迴歸分析做趨勢預測

本章將以機器學習之線性迴歸實作，以氣溫預測珍珠奶茶的銷售量，做為茶飲店用來進貨、備料的依據。



## 第 8 章

# 機器學習實戰 (一)： 用線性迴歸分析做趨勢預測

- 8-1 機器學習前準備
- 8-2 機器學習實作



# 8-1

## 機器學習前準備



資料  
科學

0

### 問個感興趣的問題



火焰黑豹珍奶  
上茶了!

- 1 使用過去幾天氣溫和銷售量的數據，以未來氣溫預測珍奶的銷售量，這樣可行嗎？
- 2 預測出來的銷售量為何？如果可以事先知道的話，對於產品的備料一定會有很大的幫助。



資料科學 × 機器學習

實戰探索

Practical Exploration





## 資料取得

- 茶飲店連續20 天所記錄的**當天氣溫**和火焰黑豹珍奶**銷售量**，先將其建置成pandas 的**資料框**。

氣溫 (°C)	銷售量 ( 杯 )
17	386
17	360
17	383
22	146
19	300
21	254
17	403
17	381
22	269
24	99

氣溫 (°C)	銷售量 ( 杯 )
21	171
21	204
21	213
17	279
25	97
21	262
20	262
19	225
19	240
22	226

```
1 import pandas as pd
2 t = [17, 17, 17, 22, 19, 21, 17, 17, 22, 24, 21, 21, 21, 17, 25, 21, 20, 19, 19, 22]
3 q = [386, 360, 383, 146, 300, 254, 403, 381, 269, 99, 171, 204, 213, 279, 97, 262, 262, 225, 240, 226]
4 df = pd.DataFrame()
5 df['T'] = t      #行資料：氣溫
6 df['Q'] = q      #行資料：銷售量
7 df.head()
```

氣溫串列 t  
銷售量串列 q

☞

	T	Q
0	17	386
1	17	360
2	17	383
3	22	146
4	19	300





資料  
科學

②

## 資料處理

- 檢查各行並適當調整資料型別
- 缺失值的補值或删除
- 删除重複值或異常值
- 將資料分割為特徵值 (Feature) 及標籤 (Label)。

((本例使用的資料很單純，不需要資料處理。))





## 01

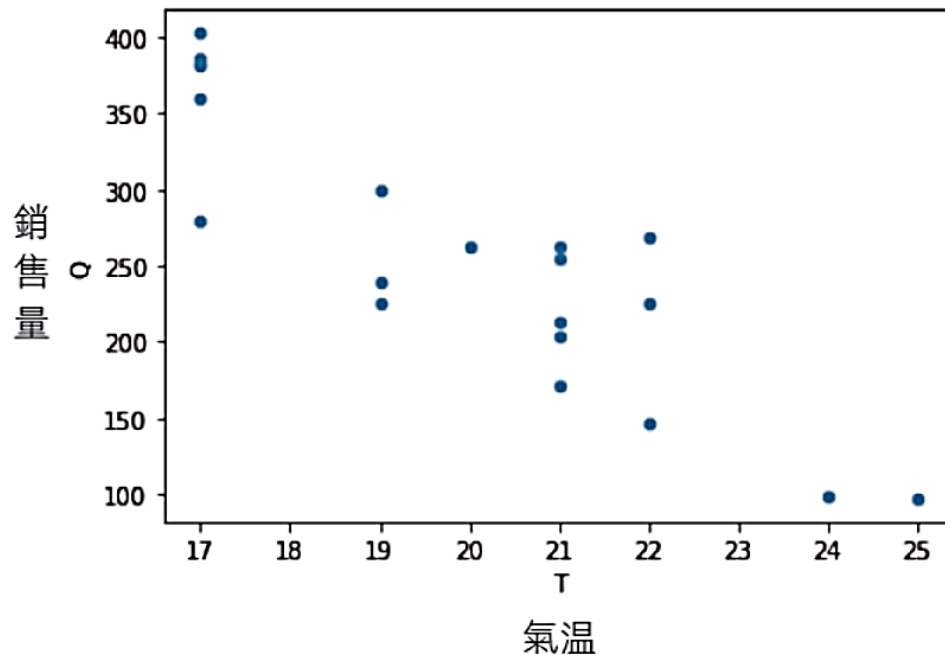
資料視覺化：由氣溫與銷售量的散佈圖中，可以看出氣溫上升，銷售量會出現遞減的情形。

氣溫高，銷售量減



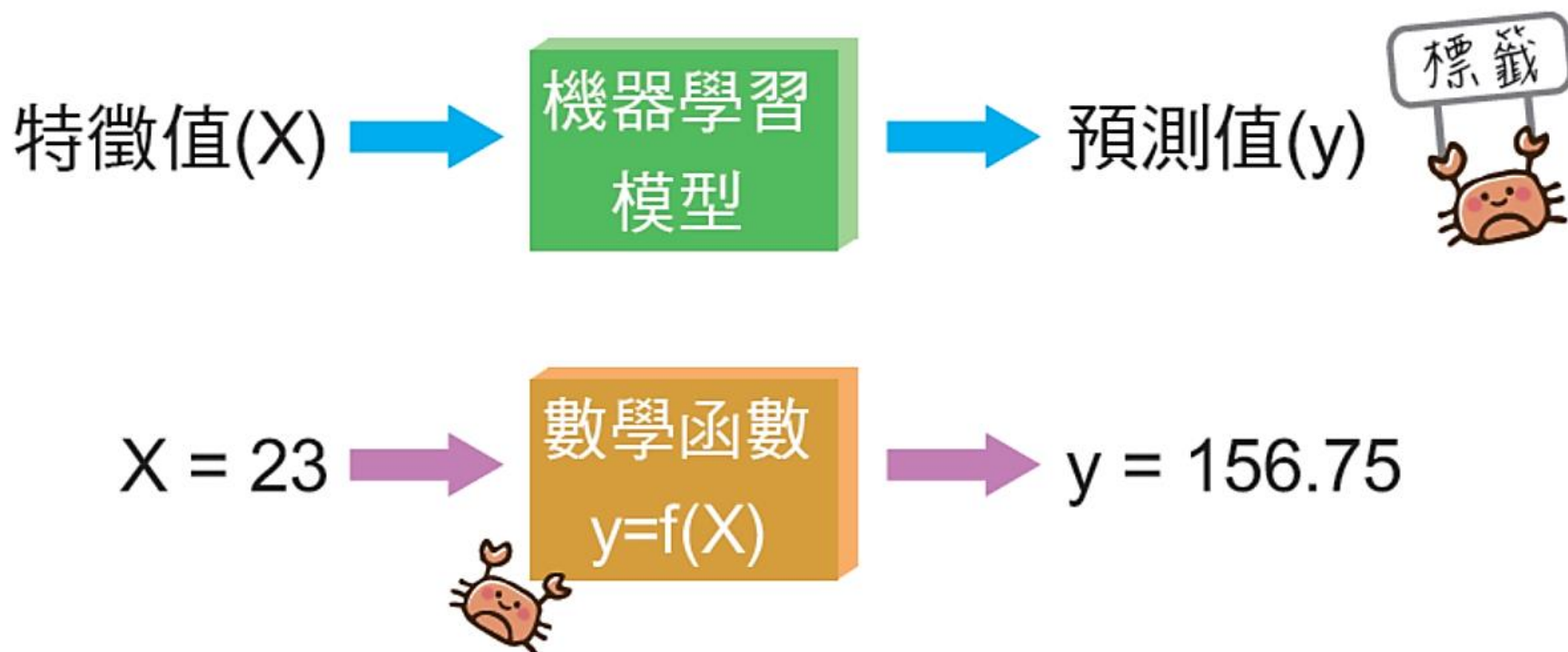
```
1 df.plot(kind='scatter', x='T', y='Q')
```

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fc05b5187f0>





- 如果要以氣溫和火焰黑豹珍奶銷售量做為機器學習中線性迴歸所需的資料，必須提供特徵值 (Feature) 以及標籤 (Label)。
- 機器學習特徵值的變數名稱慣用大寫X、標籤則是小寫y。
- 類似數學函數的概念：輸入X 值，經過運算後，得到結果y。





- 特徵值 (X) 是氣溫，而銷售量則是標籤 (y)。
- 特徵值設定成資料框 (Dataframe)。
- 標籤需取一行 (欄) 資料設定成序列 (Series)。

### 建立特徵值資料框(Dataframe)

特徵值資料框名稱 = 資料框名稱[['行索引1', '行索引2',...]]



```
df_X = df[['T']]
```

特徵值(X)

標籤(y)

氣溫 (T)

銷售量 (Q)

17	386
17	360
17	383
22	146
19	300

### 建立標籤序列(Series)

標籤序列名稱 = 資料框名稱['行索引']

```
df_y = df['Q']
```



## 02

# 建立特徵值資料框及標籤序列

- 特徵值存在「df\_X」資料框
- 標籤則存在「df\_y」序列

雙層的中括號(特徵值), 設定成資料框



```
1 df_X = df[['T']]
```

```
2 df_y = df['Q']
```

單層的中括號(標籤), 設定成序列





```
1 df_X.head()
```

印特徵值出來看看



T

0 17

1 17

2 17

3 22

4 19





```
1 df_y.head()
```

印標籤出來看看



```
0      386
```

```
1      360
```

```
2      383
```

```
3      146
```

```
4      300
```

```
Name: Q, dtype: int64
```



## 8-2

# 機器學習實作

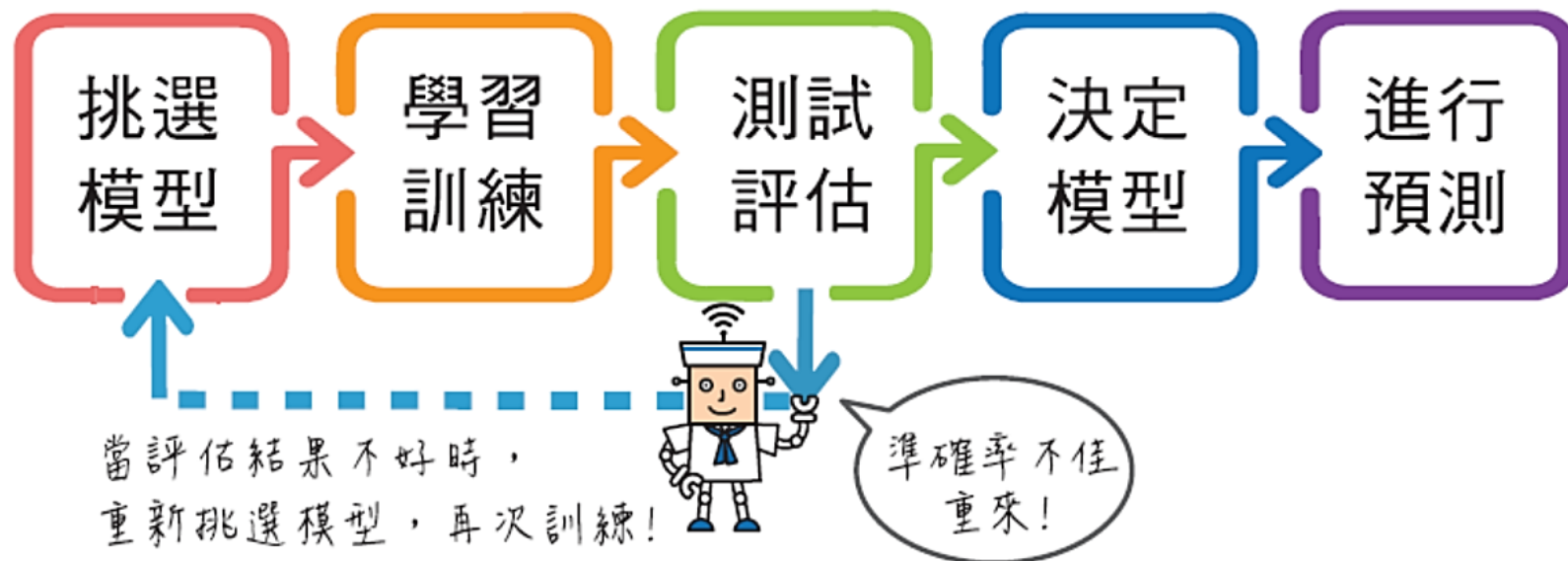


資料  
科學

4

## 機器學習做資料分析

### 機器學習實作步驟



## 8-2-1 提出具體的假設

- 由前面的「步驟3 探索性資料分析」得知「**氣溫上升，銷售量遞減**」。
- 提出以下假設：「**由機器利用氣溫為特徵值，自行學習如何準確預估銷售量**」。
- 經過機器學習的過程，再輸入未來的氣溫到訓練產生的模型中，進行銷售量的趨勢預測。



## 8-2-2 找出機器學習模型



①

### 挑選模型：匯入線性迴歸模型

首先匯入 `sklearn` 中的線性迴歸模型。



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







## 學習訓練：建立並訓練線性迴歸模型

建立模型

`LinearRegression()`

訓練模型

`fit()`

建立線性迴歸模型

線性迴歸模型名稱 = `LinearRegression()`

訓練線性迴歸模型

線性迴歸模型名稱.`fit`(訓練特徵值資料框名稱, 訓練標籤序列名稱)



- 有了包含**特徵值** (即氣溫df\_X) 和**標籤** (即銷售量df\_y) 的**訓練用資料**，接下來就以此做為模型訓練的輸入。

```
1 lm = LinearRegression()  
2 lm.fit(df_X, df_y)
```

建立新模型 lm

訓練模型

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



### ③ 測試評估

此例資料量很少，省略「測試評估」步驟。



機器  
學習

4

## 決定模型：取出線性迴歸模型的 $m$ 、 $b$ 參數

產生模型

.coef\_  
.intercept\_

取出線性迴歸模型  $m$  參數

線性迴歸模型名稱.coef\_

取出線性迴歸模型  $b$  參數

線性迴歸模型名稱.intercept\_

$$y=f(x)=mx+b$$



資料科學 × 機器學習

實戰探索

Practical Exploration

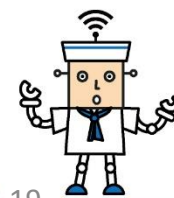


- 挑選的模型經過訓練後可以得到模型的**參數**。
- 在本例中所挑選的線性迴歸模型是創建一條描述訓練資料的**最適直線**  $y = mx + b$ 。
- 得到的**參數**  $m$  約為-33.2、**參數**  $b$  約為920.28。
- 訓練後的模型即為  $y = -33.2x + 920.28$ 。

```
1 print('線性迴歸的模型為  $y = f(x) = mx + b$ ')  
2 print('m 為 ', lm.coef_)  
3 print('b 為 ', lm.intercept_)
```

取出  $m$  跟  $b$

```
↳ 線性迴歸的模型為  $y = f(x) = mx + b$   
m 為 [-33.19704219]  
b 為 920.2809917355372
```





## 進行預測

進行預測  
predict()

線性迴歸模型預測

模型名稱.predict(特徵值資料框名稱)



- 創建模型後就可以拿來試試其準確率！
- 取一個或多個未來可能發生的**特徵**（即氣溫）當成**預測函式predict()**的輸入。
- 此處因為要做**結果預測**（即可能的銷售量），所以就**不需要標籤**（即實際的銷售量）。





```
1 temp = [[23]]  
2 p = lm.predict(temp)  
3 print(p)
```

輸入一個特徵值 (氣溫) 讓模型預測

輸出預測的銷售量

☞ [156.74902131]



```
1 temp = [[23],[18],[36]]  
2 p = lm.predict(temp)  
3 print(p)
```

輸入多個特徵值 (氣溫) 讓模型預測

☞ [ 156.74902131 322.73423227 -274.81252719]



## 8-2-3 提出機器學習後的資料分析結果

- 以簡單的**數學線性迴歸**做為**機器學習演算法**，輸入歷史資料進行訓練，利用完成的模型進行趨勢預測的確是可行的。
- 本例中，輸入一筆**特徵值資料** **[[23]]** (即氣溫 23°C)，預測出來的銷售量約為157 杯！
- 本例以氣溫的變化來預測火焰黑豹珍奶可能的銷售量。藉由預測出來的銷售量，就可以做為老闆採購備料的重要參考。







## 加廣 知識

這樣的機器學習可以進一步應用於以下情境：

- 投入廣告費與銷售額。
- 最高氣溫與尖峰用電量。
- 由身高預測體重多少。

