

Day 9

Linear Regression

[全民瘋AI系列]



第12屆 iT邦幫忙 鐵人賽

Day 9 學習目標

01

認識線性回歸

透過機器學習來找出一條函式，來最佳化模型

02

線性回歸程式手把手

簡單線性回歸、多元回歸、非線性回歸

Part 1

線性迴歸 (Linear Regression)

觀念講解



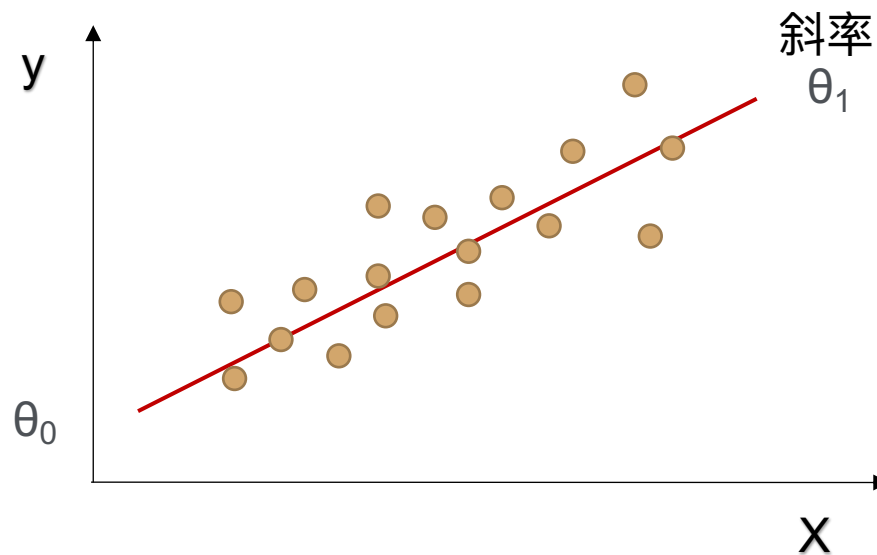
第12屆 iT邦幫忙 鐵人賽

/// 線性迴歸(Linear Regression)

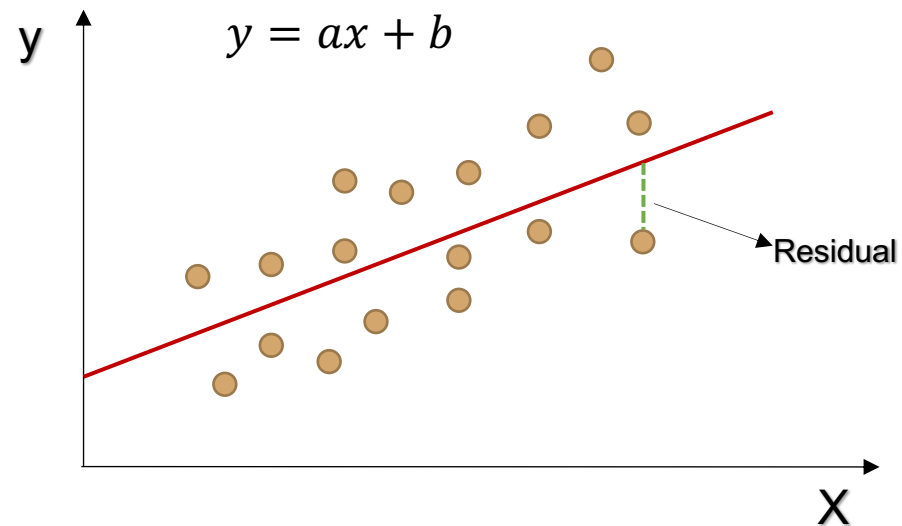
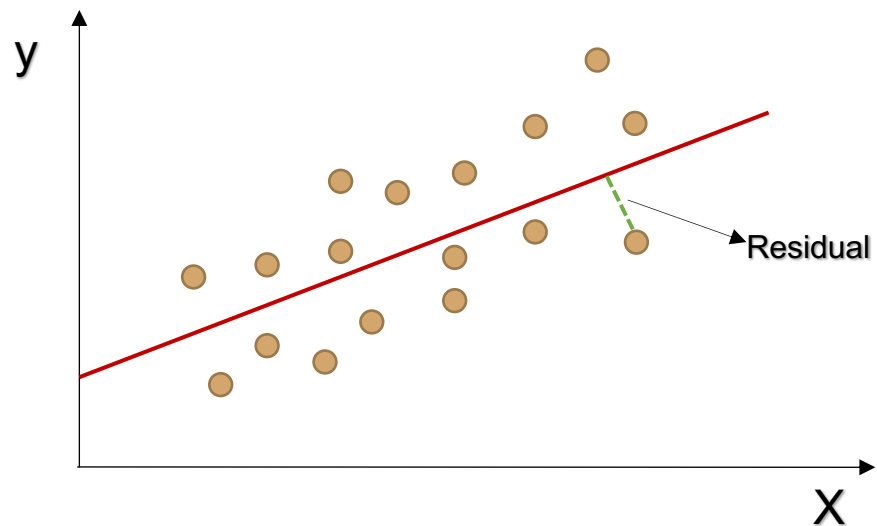
- 簡單線性回歸: $y = \beta_0 + \beta_1 x$
- β_0 截距(Intercept)
- β_1 斜率(Slope)
- x 為輸入， y 輸出

透過機器學習(ML)來找出一條函式，來最佳化模型

目標 = 降低 Error



哪一個是誤差的測量方式呢？



$$\text{目標函式}(\theta_0, \theta_1) = \frac{1}{2n} \sum (\hat{y}_i - y_i)^2$$

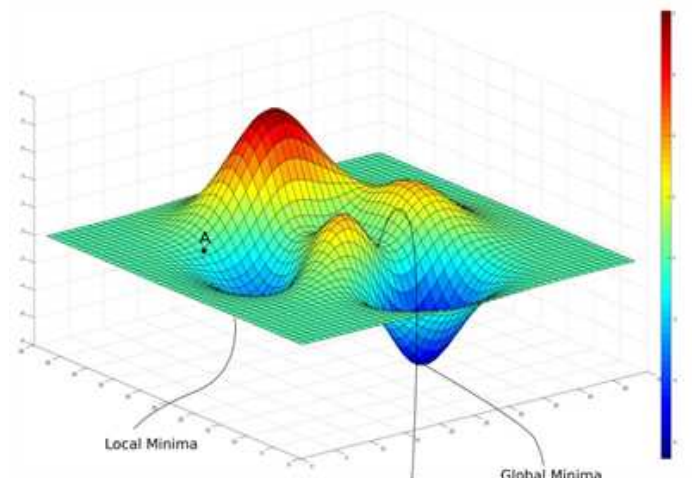


兩種求解方法

- Close form

$$\theta = (X^T X)^{-1} X^T y$$

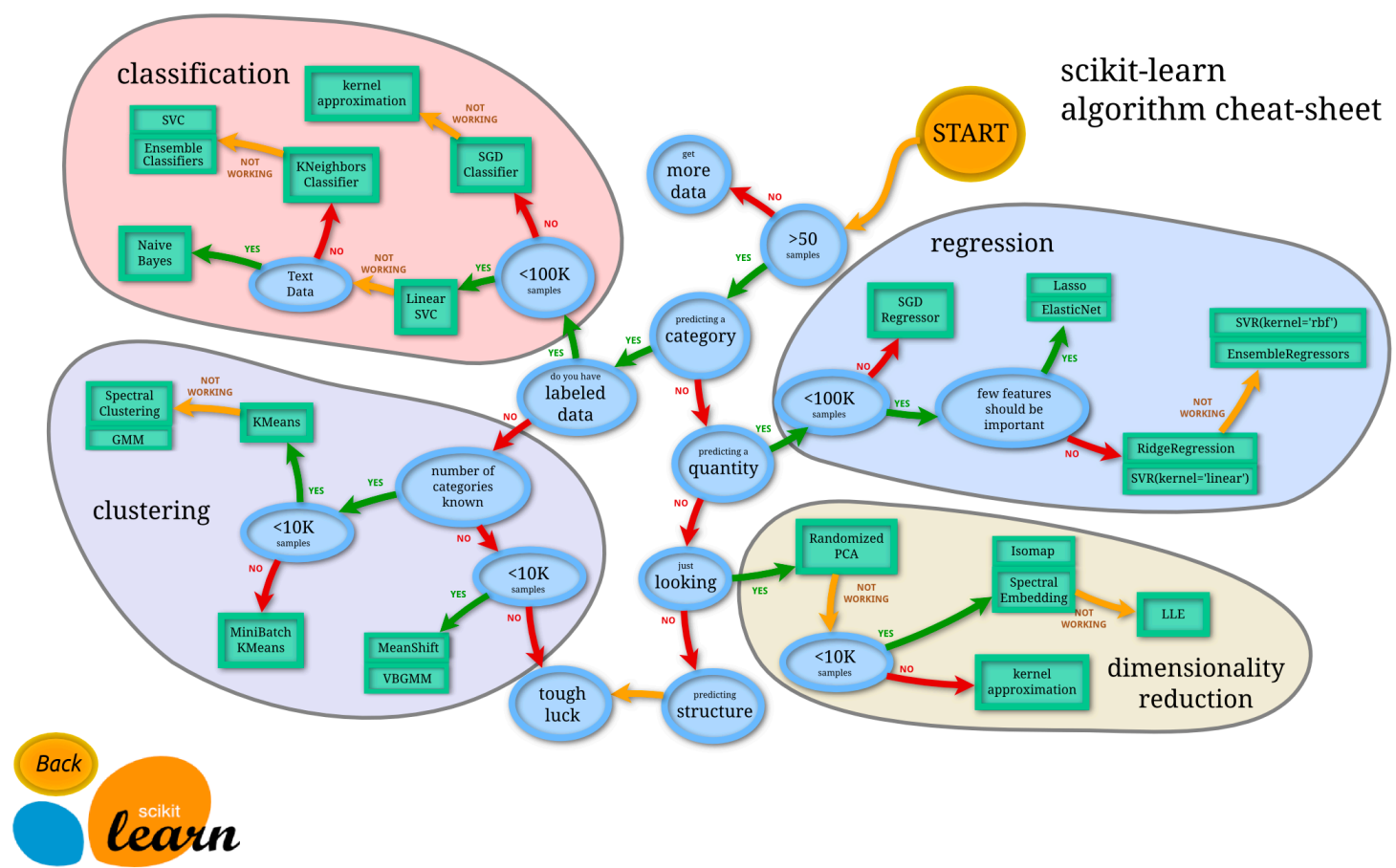
- Gradient descent



[台大 李宏毅教授 ML Lecture 3-1: Gradient Descent](#)



sklearn 在 Python 中提供大量常見的機器學習演算法和許多實用的資料集合，其中今天就是要使用到它的現性回歸的演算法 (Linear Regression)。

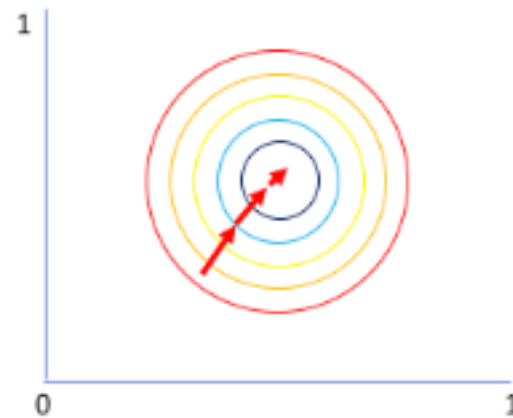
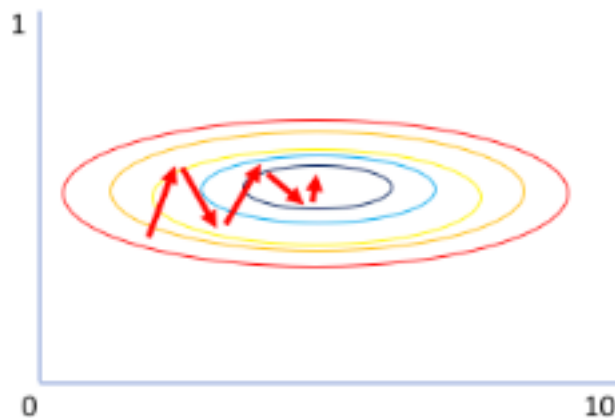


Linear Regression

- Linear Regression 使用時機
 - 標籤(y)為連續值
 - 資料量較少 (<100k)
 - 假設資料特徵和標籤之間有線性關性

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3$$

- 透過Feature Transformation(特徵轉換)、 Normalization(正規化) 幫助快速收斂



//// scikit-learn 模型評估方法

- Regression
 - Mean Absolute Error
 - Root Mean Absolute Error
 - R2 Score
- Classification
 - Accuracy
 - F1 Score

$$\text{MSE} = \frac{1}{M} \sum_{i=1}^M (x_i - y_i)^2$$

Diagram illustrating the Mean Squared Error (MSE) formula:

- M : Number of data
- x_i : Predicted output
- y_i : Desired output

```
from sklearn import metrics  
  
mae = metrics.mean_absolute_error(y, prediction)  
mse = metrics.mean_squared_error(y, prediction)  
r2 = metrics.r2_score(y, prediction)
```



Part 2

線性迴歸 (Linear Regression) 程式實作

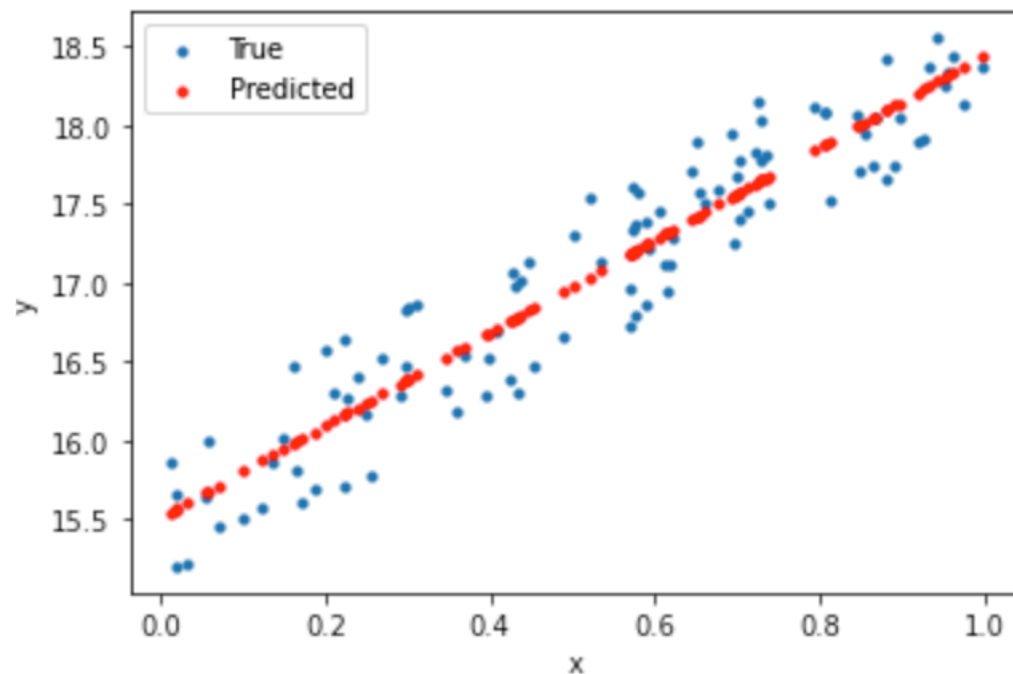


第12屆 iT邦幫忙 鐵人賽

//// 簡單線性回歸

- 目標函式: $y=3x+15$
- 隨機添加 noise 讓資料分散
- x 值域介於 0~1

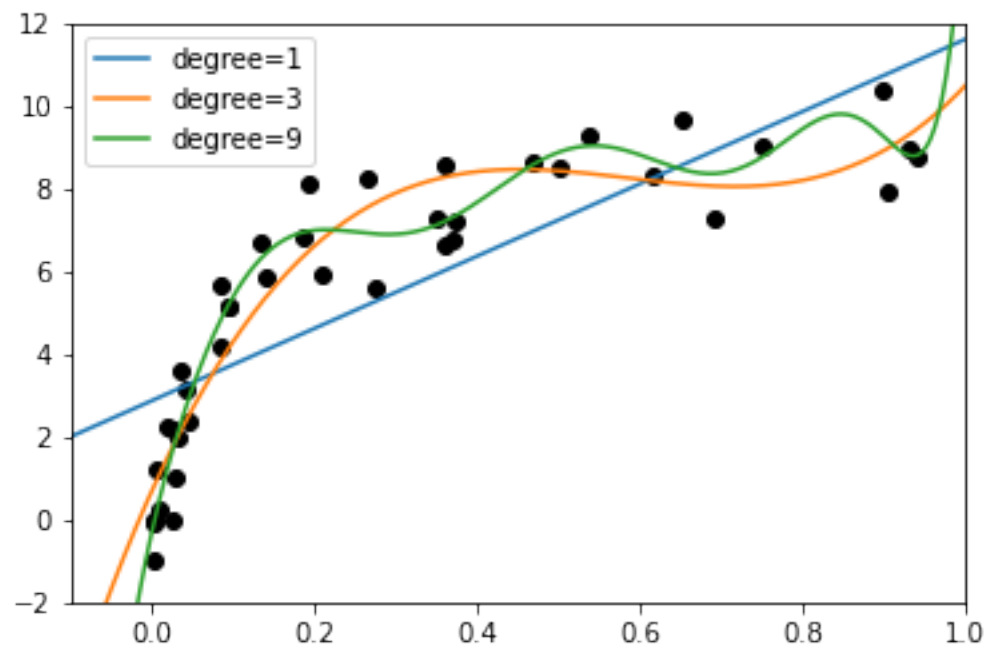
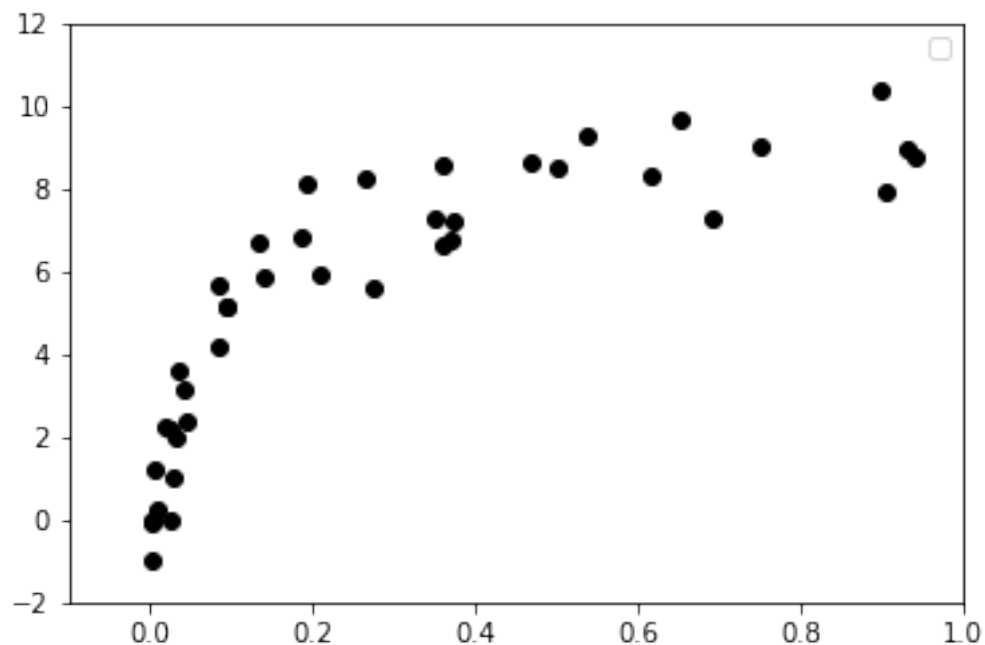
```
from sklearn.linear_model import LinearRegression  
  
# 建立LinearRegression模型  
linearMmodel = LinearRegression(fit_intercept=True)  
# 使用訓練資料訓練模型  
linearMmodel.fit(x, y)  
# 使用訓練資料預測  
predicted = linearMmodel.predict(x)
```



//// [非線性回歸] 訓練一個多項式模型

在 sklearn 中如果要訓練一個多項式的回歸模型：

- PolynomialFeatures 生成多項式特徵
- LinearRegression



Thanks

PRESENTED BY 10程式中



第12屆 iT邦幫忙 鐵人賽