

## 第1章 機器學習之線性回歸



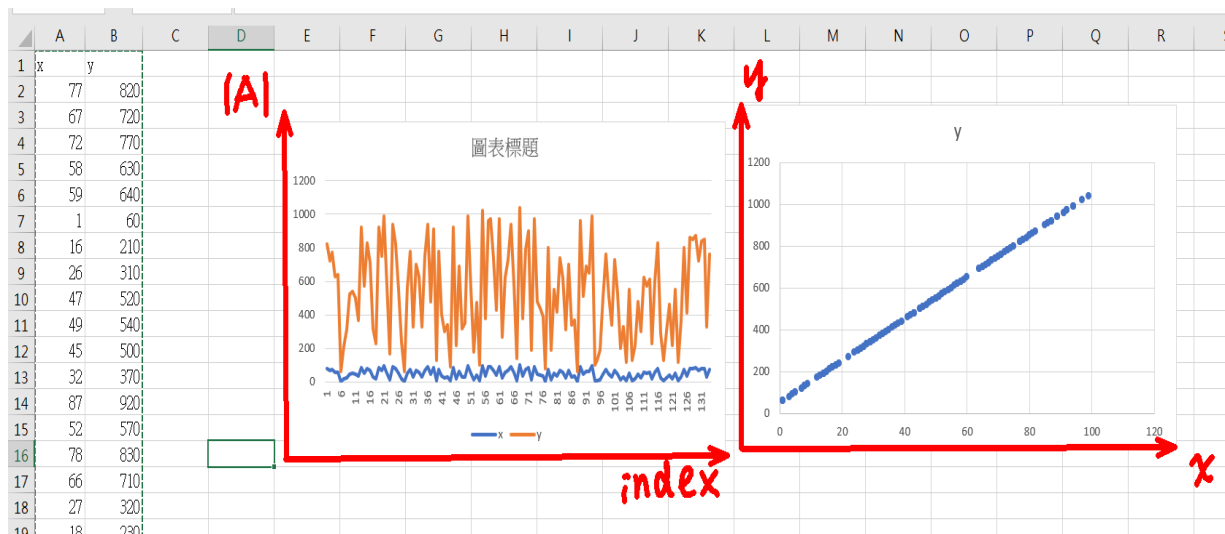
**章節目標：**在這一章節，我們介紹什麼是線性回歸，並用幾個例子說明線性回歸的應用，先用Excel操作，在用Python 來實現，同時熟悉一下follow **CRISP-DM** (跨領域資料分析作業流程) 方法論來建立ML 模型。



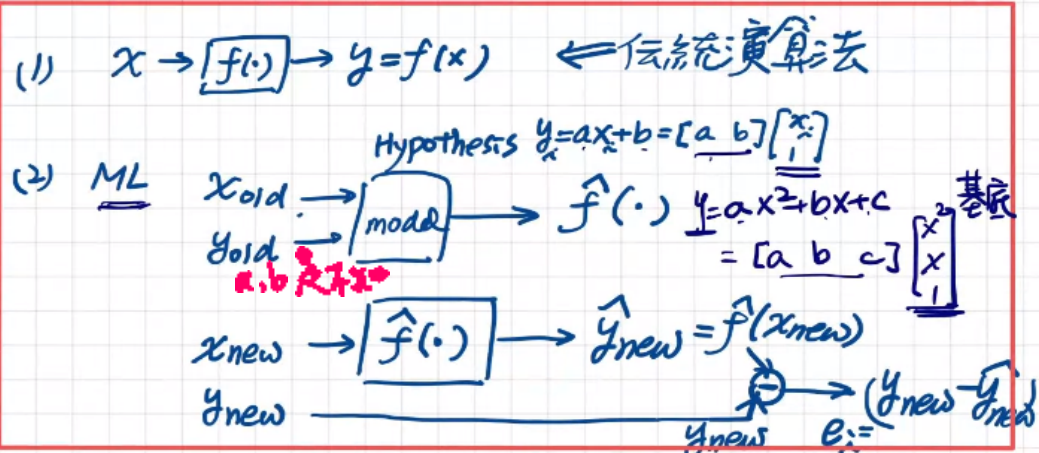
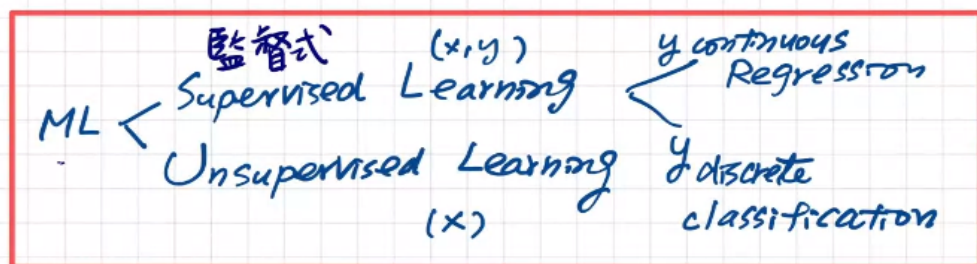
本章學習  
內容摘要

1. **EXCEL** 體驗線性迴歸
2. 線性迴歸的數學理論
3. 梯度下降法
4. **COLAB** 環境
5. **PYTHON** 實作簡單線性回歸

## 1. 線性迴歸EXCEL 體驗



## 2. 線性回歸的數學理論



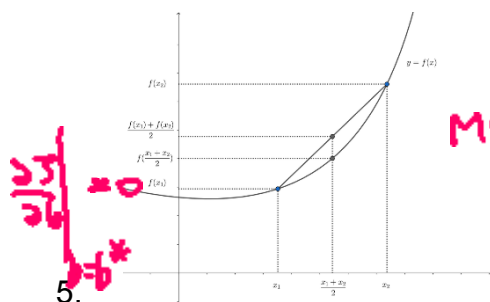
meet.google.com 正在共用你的畫面。

## 略

Given  $(x_i, y_i)$   
 $i=1 \sim 100 \Rightarrow$  Error function  

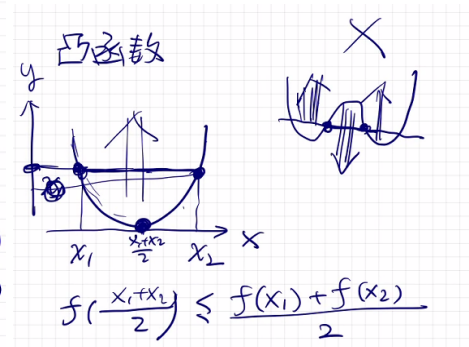
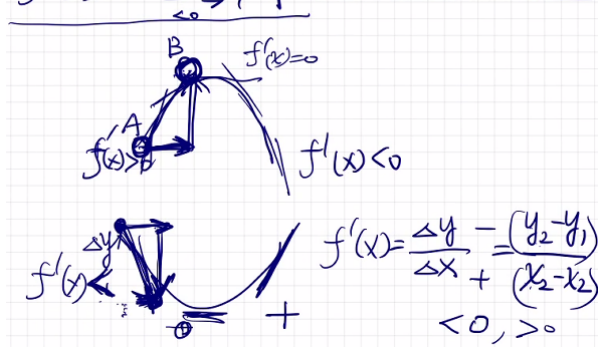
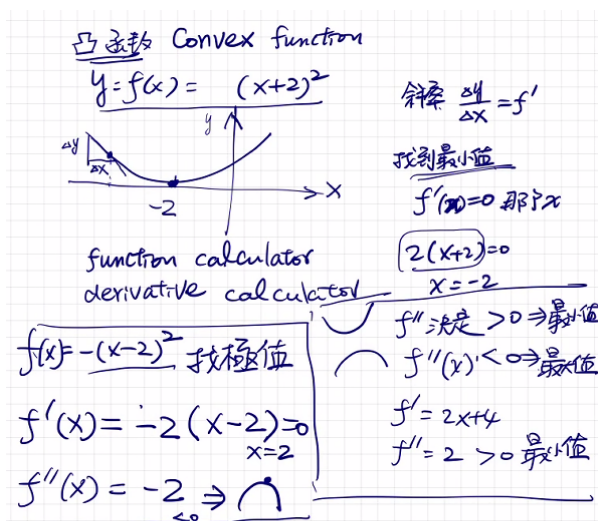
$$\sum_{i=1}^n \left( y_i - (ax_i + b) \right)^2$$

对  $a, b$  来说  $f$  是  
convex function

$$\frac{\partial \mathcal{L}}{\partial \alpha} = 0$$


MSE





## 6. 延伸參考

### 1. 台大李宏毅 線性回歸

[http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML\\_2017/Lecture/Regression.pdf](http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2017/Lecture/Regression.pdf)

### 2. 白板推导系列 P9(系列三) 线性回归1-最小二乘法及

<https://www.bilibili.com/video/BV1aE411o7qd?p=9>

作業: 建議速度 30字↑/min

<https://www.typing.com/student/typing-test/1-minute>

函數網站

## 7. 實作

[Colab \(Google\)](#)

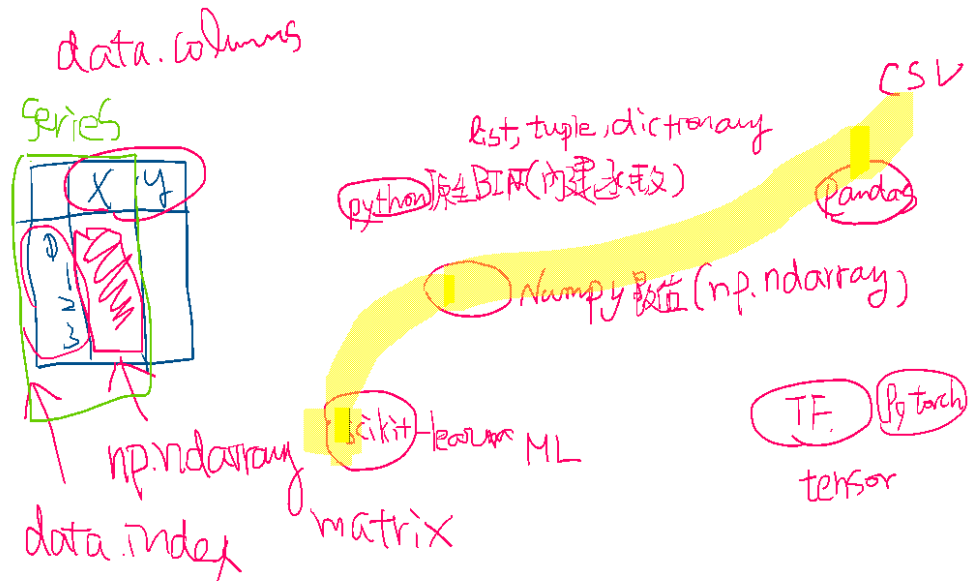
Text	Note
<p># L3 Example 1 : Linear Regression</p> <p>We will follow CRISP-DM design method</p> <p>&lt;img src="https://www.symbolab.com/public/image s/home_left.png" width="300" height="300"&gt;</p> <p>Step 1: Load data</p>	<p>HTML語法 :</p> <p>&lt;img&gt;插入圖片</p> <p>可去看W3Cschool</p> <pre>import numpy as np import matplotlib.pyplot as plt import pandas as pd  data= pd.read_csv("你的csv檔名") data.head()</pre>
<p>Step 2: Prepatre X, Y</p>	<pre>X=data.x.values.reshape(-1,1) Y=data.y.values.reshape(-1,1) print(X.shape)</pre>
<p>Step 3: Build ML Model</p>	<pre># second choice from sklearn import linear_model model = linear_model.LinearRegression() model.fit(X,Y)  a=model.coef_ b=model.intercept_  preY=model.predict(X)</pre>

	<pre>data['preY']= preY data.head()</pre>
Step 4: Evaluate Model	<pre>from sklearn.metrics import r2_score from sklearn.metrics import mean_squared_error as MSE from sklearn.metrics import mean_absolute_error as MAE  r2score=r2_score(Y, preY) mse=MSE(Y, preY) mae=MAE(Y, preY) print("R2score=",r2score) print("MSE=",mse) print("MAE=",mae)</pre>
Step 5: Export Model (Write out result, deploy model)	

### 1. Review 上次內容

- 線性回歸,我們介紹什麼是Excel操作並用幾個例子說明線性回歸的應用, 感測器校正
- 在用Python 來實現=>google Coble
- ML 實作方法熟悉一下follow **CRISP-DM** (跨領域資料分析作業流程) 方法論來建立ML 模型
  - Step 1: import library, load data
  - Step 2: prepare “feature”
    - All-in今天多元回歸 (multilinear regression) 會著重select feature 方法,
    - 還有包括一些前處理 (missing data=> imputation, normalization, data type 例如 categorical data onehot encoding, Label\_Encoding,

- 資料型別的 transformation) 各Library 之間的資料溝通都是透過 Numpy (np.ndarray)



- Step 3: Build model
- Step 4: Evaluate model
  - ◆ MSE, MAE, R2

```
from sklearn.metrics import r2_score as R2
from sklearn.metrics import mean_squared_error as MSE
yPre=model.predict(X)
print("MSE=",MSE(Y,yPre))
print("R2=",R2(Y,yPre))
```



byam.github.io/dlnd/2017/11/13/model-evaluation-and-validation.html

- What is the simplest possible model that fits a bunch of points?
  - the **average of all the values** and **draw a horizontal line** through them.

R2 SCORE

**BAD MODEL**

The errors should be similar.  
R2 score should be close to 0.

**GOOD MODEL**

The mean squared error for the linear regression model should be a lot smaller than the mean squared error for the simple model.  
R2 score should be close to 1.

$R^2 = 1 - \frac{\text{Area of pink squares}}{\text{Area of green squares}}$

### ◆ Performance Metrics MSE VS. R2

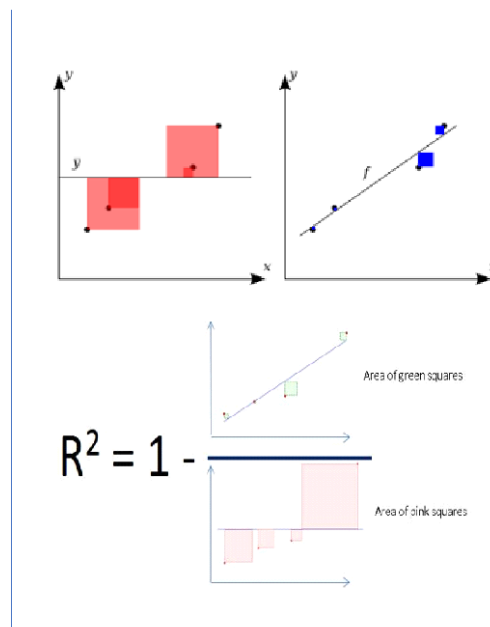
R2 square

<https://byam.github.io/dlnd/2017/11/13/model-evaluation-and-validation.html>

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i^*)^2 = J$$

$\hat{y}_i^* = a^* \cdot x_i + b^*$

$$a^*, b^* = \arg \min_{a, b} J$$





## 1. MACHINE LEARNING 開發工具

- (1). Colab  $\Rightarrow$  online ML, DL, ...開發工具

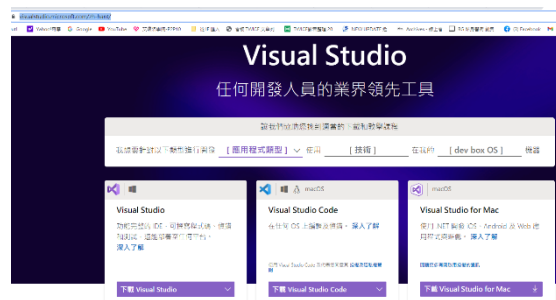
<https://colab.research.google.com>

- (2). Anaconda  $\Rightarrow$  python, IDE

<https://www.anaconda.com/products/individual>

- (3). Visual Studio  $\Rightarrow$  general IDE

[visualstudio.microsoft.com](https://visualstudio.microsoft.com)



- (4). Sublime Text 3 (good editor) $\Rightarrow$  尤其是開發 frontend PHP, HTML web

<https://www.sublimetext.com/3>

- (5). Eclipse (Sun) Java  $\Rightarrow$  general IDE

- (6). HeidiSQL  $\Rightarrow$  編寫資料庫

<https://www.heidisql.com/>

HFS

<https://www.facebook.com/freewarefans/posts/10152105749348487/>

<https://www.azofreeware.com/2006/03/hfs-20.html>

本頁勿刪

