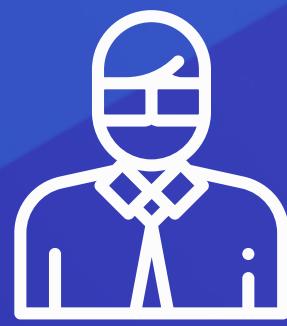


Day 35

機器學習

# Regression vs. Classification



楊証琨

出題教練

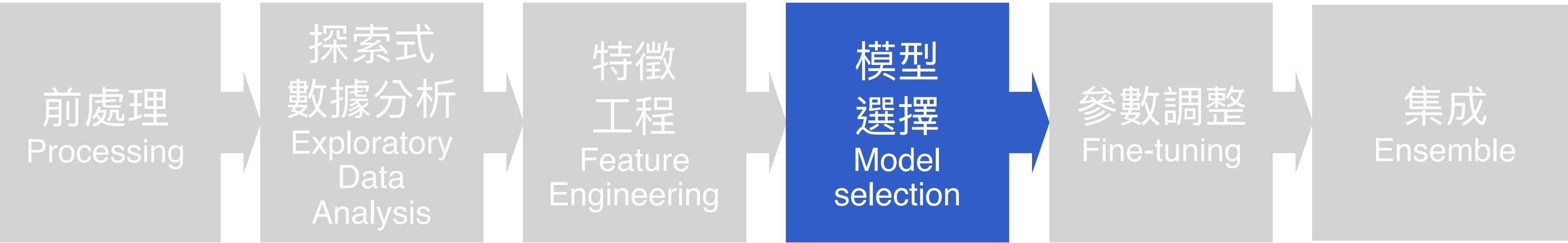


# 知識地圖 機器學習- 模型選擇 - Regression vs. Classification



## 機器學習基礎模型建立

### 監督式學習 Supervised Learning



### 非監督式學習 Unsupervised Learning



## 模型選擇 Model selection

### 概論

驗證基礎

預測類型

評估指標

### 基礎模型 Basic Model

線性回歸 Linear Regression

邏輯斯回歸 Logistic Regression

套索算法 LASSO

嶺回歸 Ridge Regression

### 樹狀模型 Tree based Model

決策樹 Decision Tree

隨機森林 Logistic Regression

梯度提升機 Gradient Boosting Machine

# 本日知識點目標

- 了解機器學習中迴歸與分類的差異
- 回歸問題與分類問題的定義
- 什麼是多分類問題？多標籤問題？

# 回歸 vs. 分類

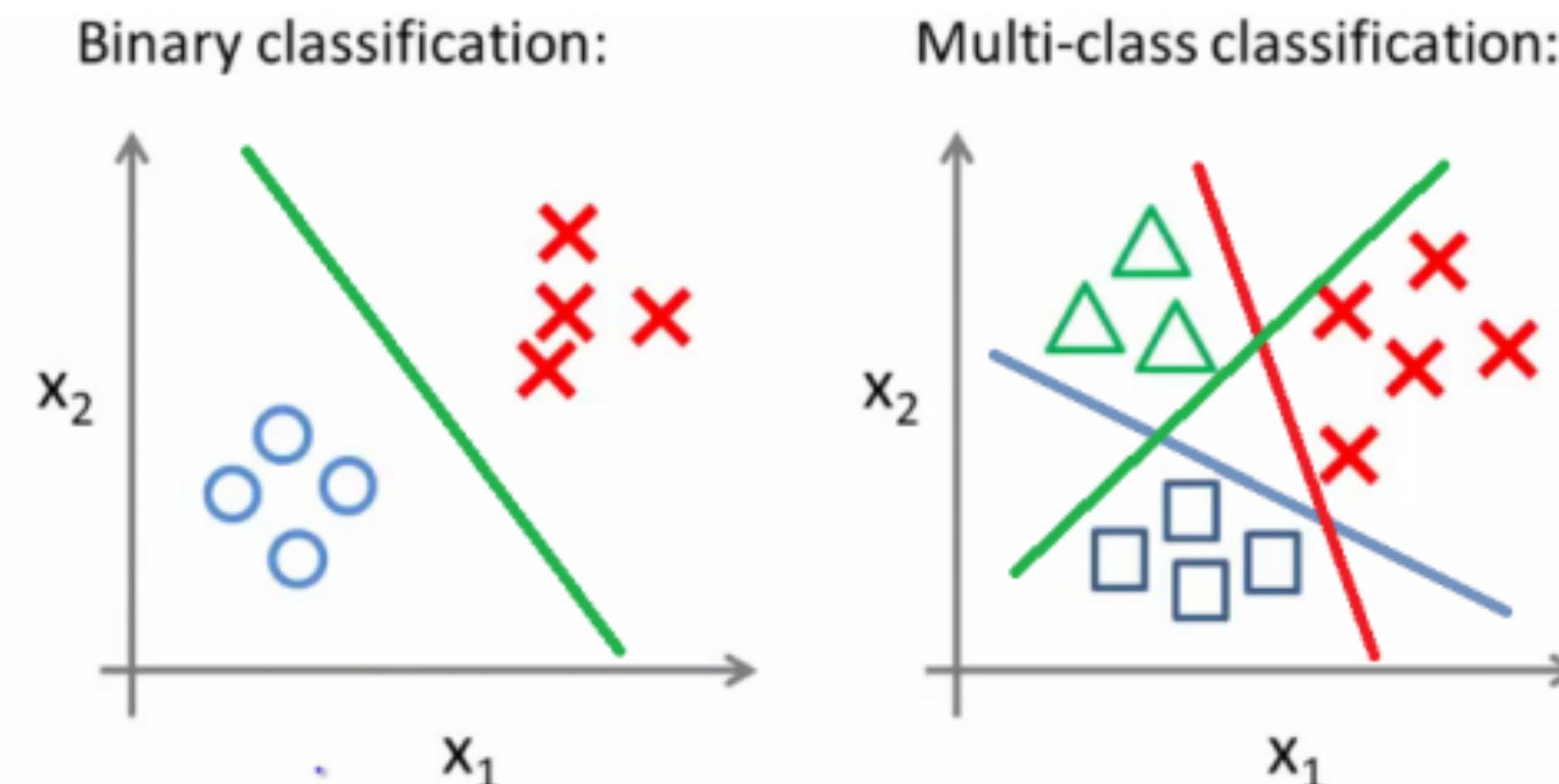
- 機器學習的監督式學習中主要分為回歸問題與分類問題。
- 回歸代表預測的目標值為實數 ( $-\infty$  至  $\infty$ )
- 分類代表預測的目標值為類別 (0 或 1)

# 回歸問題可轉換

- 回歸問題是可以轉化為分類問題
  - 模型原本是直接預測身高 (cm)
  - 可更改為預測是否高於中位數 (yes or no)
  - 或是預測多個類別，如矮、中等、高
- 可根據專案的需求自行變化目標定義

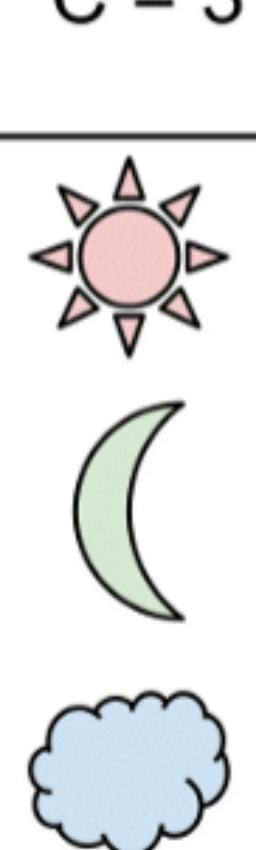
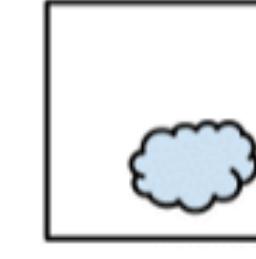
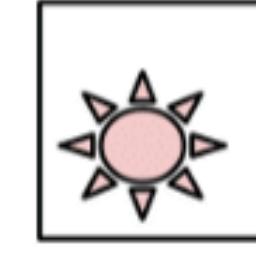
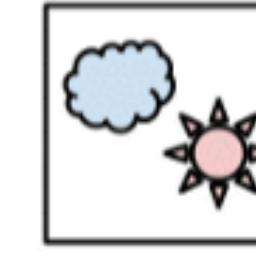
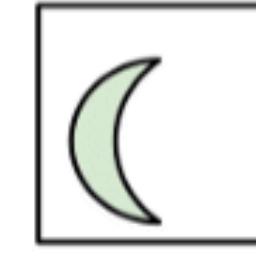
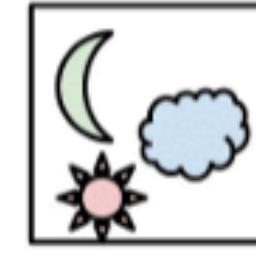
# 二元分類 (binary-class) vs. 多元分類 (Multi-class)

- 二元分類，顧名思義就是目標的類別僅有兩個。像是詐騙分析 (詐騙用戶 vs. 正常用戶)、瑕疵偵測 (瑕疵 vs. 正常)
- 多元分類則是目標類別有兩種以上。如手寫數字辨識有 10 個類別 (0~9)，影像競賽 ImageNet 更是有高達 1,000 個類別需要分類



# Multi-class vs. Multi-label

- 當每個樣本都只能歸在一個類別，我們稱之為多分類 (Multi-class) 問題；而一個樣本如果可以同時有多個類別，則稱為多標籤 (Multi-label)。
- 了解專案的目標是甚麼樣的分類問題並選用適當的模型訓練。

	Multi-Class	Multi-Label
C = 3	Samples	Samples
	  	  
	Labels (t)	Labels (t)
	[0 0 1]   [1 0 0]   [0 1 0]	[1 0 1]   [0 1 0]   [1 1 1]



# 延伸 閱讀

- 了解回歸與分類的差異在哪裡?  
網頁連結：[回歸與分類的比較](#)

Property	Supervised Classification	Regression
output type	discrete(class labels)	continuous(number)
What are you trying to find?	decision boundary	"best fit line"
Evaluation	accuracy	Risidual sum of square(RSS)



解題時間

It's Your Turn

請跳出PDF至官網Sample Code & 作業  
開始解題

