

# 機器學習

# 梯度提升機



## 出題教練

楊証琨

# 知識地圖 機器學習- 模型選擇 - 樹狀模型 - 梯度提升機(Gradient Boosting Machine)



## 機器學習概論 Introduction of Machine Learning

### 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 Random Forest

梯度提升機 Gradient Boosting Machine



# 本日知識點目標

- 了解梯度提升機的基本原理與架構
- 梯度提升機與決策樹/隨機森林的差異
- 梯度提升機的梯度從何而來，又是怎麼計算的

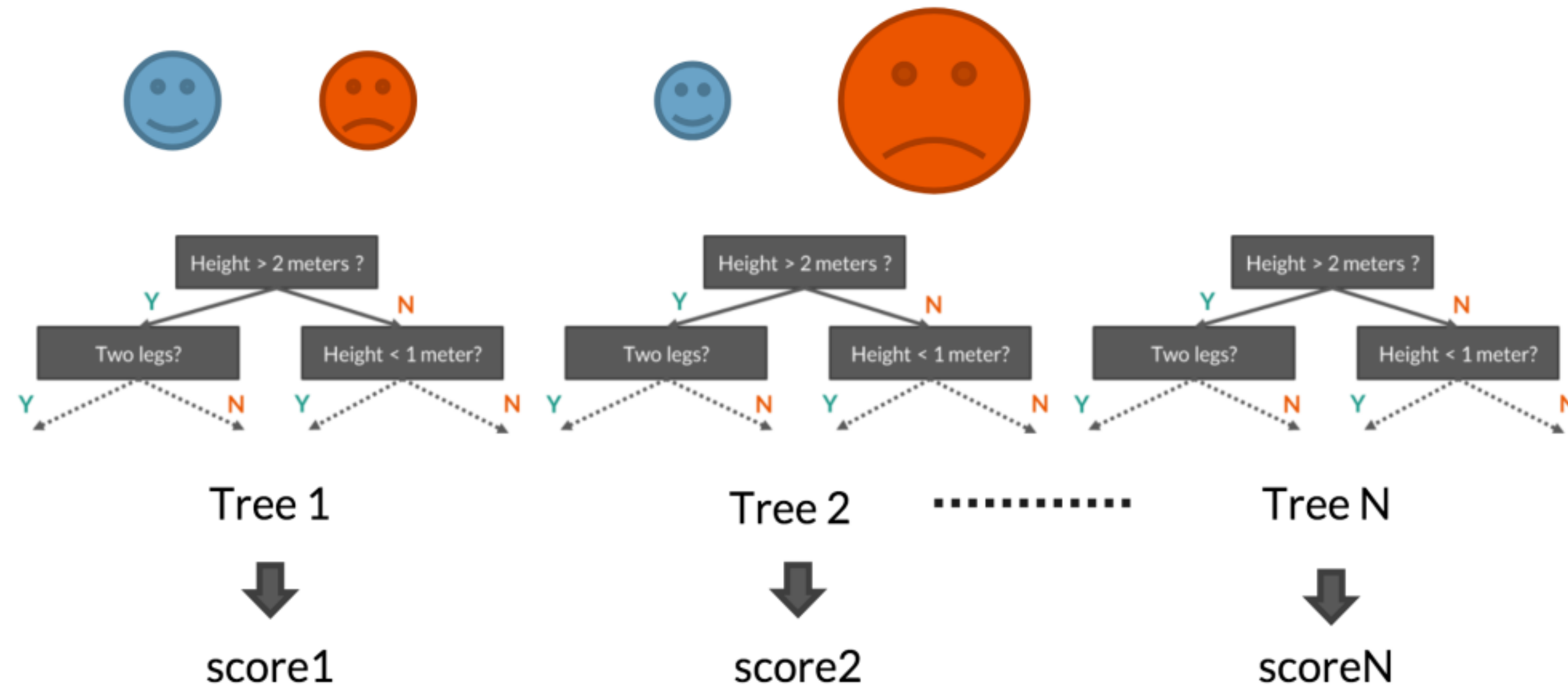
# 梯度提升機 (Gradient Boosting Machine)

---

- 隨機森林使用的集成方法稱為 Bagging (Bootstrap aggregating)，用抽樣的資料與 features 生成每一棵樹，最後再取平均
- Boosting 則是另一種集成方法，希望能夠由後面生成的樹，來修正前面樹學不好的地方
- 要怎麼修正前面學錯的地方呢？計算 Gradient!

# 梯度提升機 (Gradient Boosting Machine)

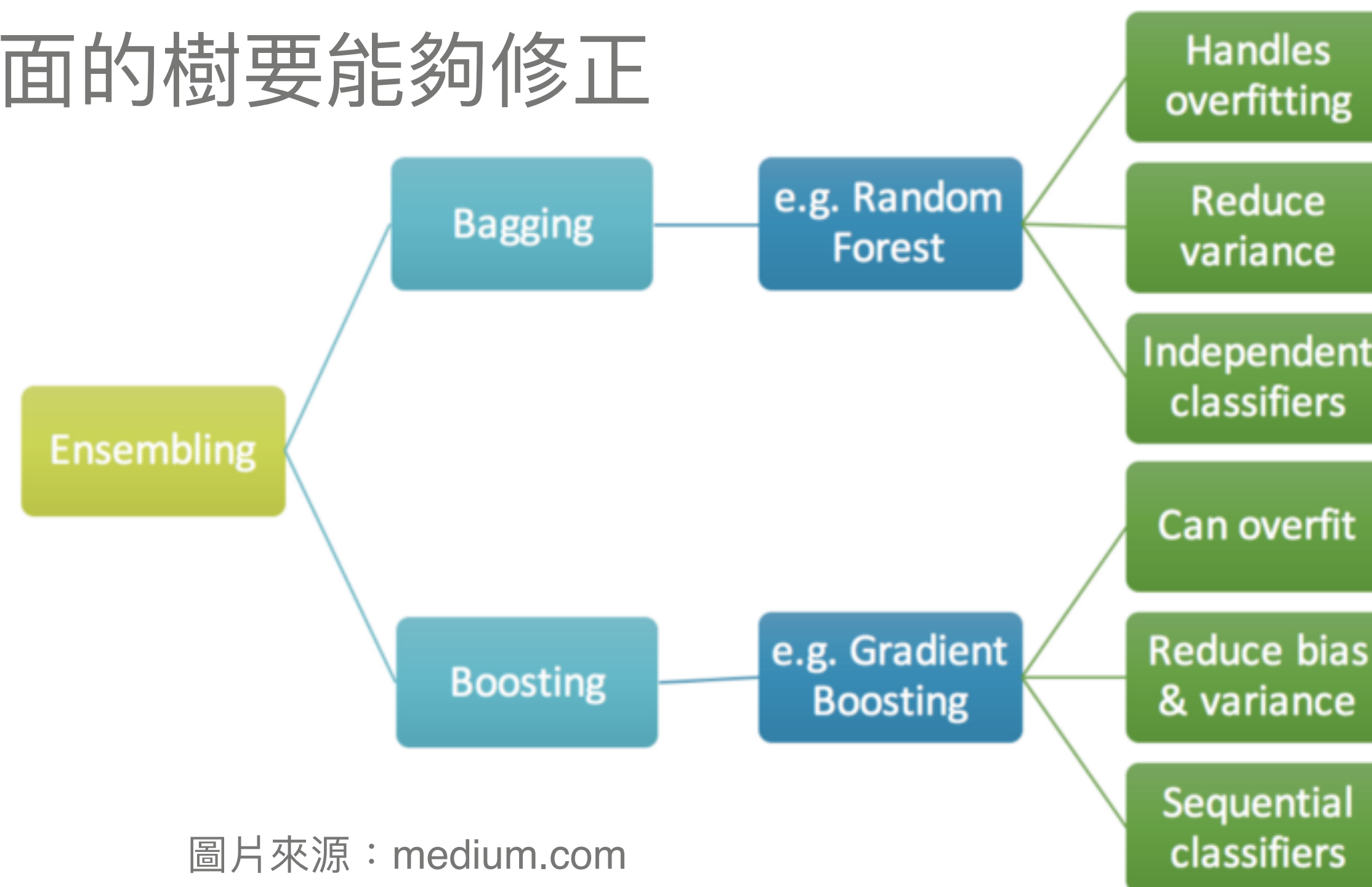
- 每次生成樹都是要修正前面樹預測的錯誤，並乘上 learning rate 讓後面的樹能有更多學習的空間



$$y = \text{score1} * \text{learning\_rate} + \text{score2} * \text{learning\_rate} + \dots + \text{scoreN} * \text{learning\_rate}$$

# Bagging 與 Boosting 的差別

- Bagging 是透過**抽樣 (sampling)** 的方式來生成每一棵樹，樹與樹之間是獨立生成的
- Boosting 是透過**序列 (additive)** 的方式來生成每一顆樹，每棵樹都會與前面的樹關聯，因為後面的樹要能夠修正





# 充電時間 Brain Charge

請跳出PDF至官網Sample Code & 作業  
進行今日作業

