

Day 45

機器學習

梯度提升機

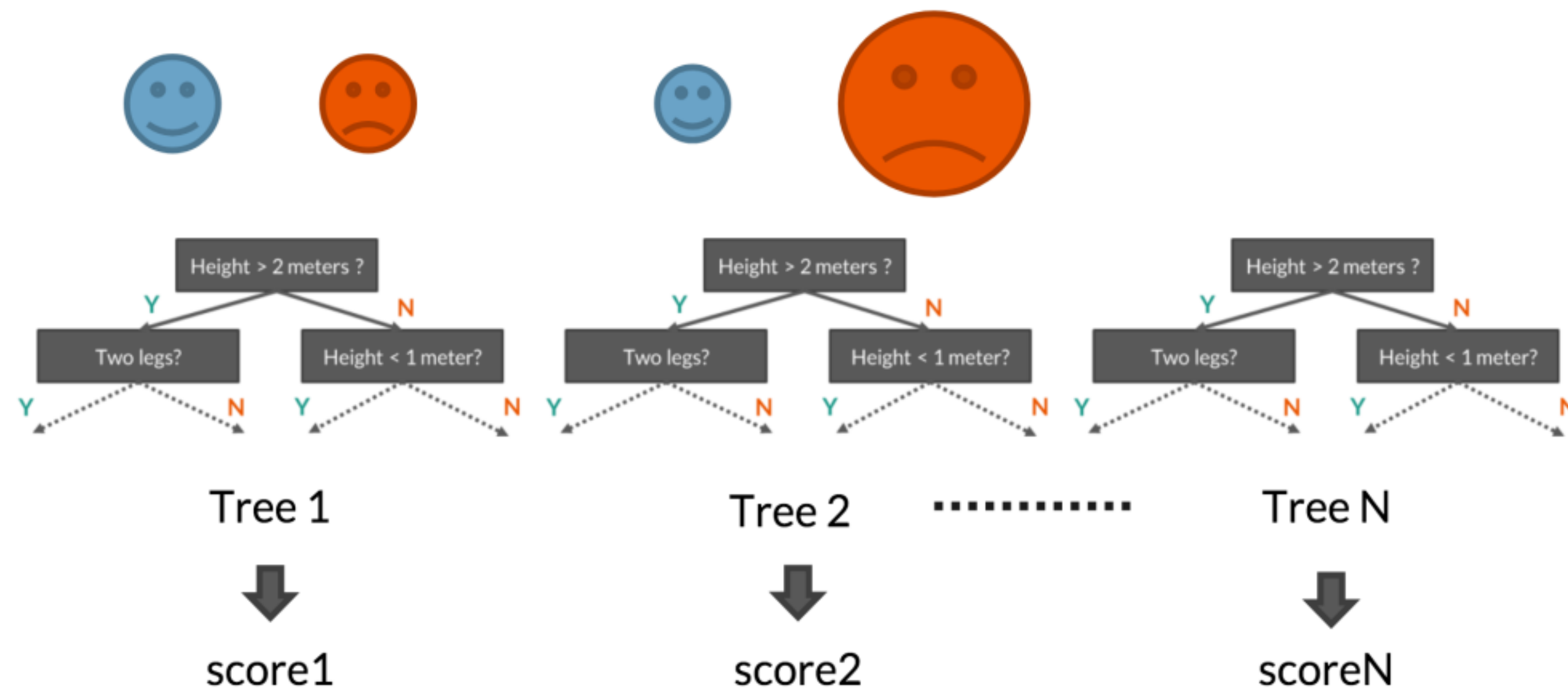


梯度提升機 (Gradient Boosting Machine)

- 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)

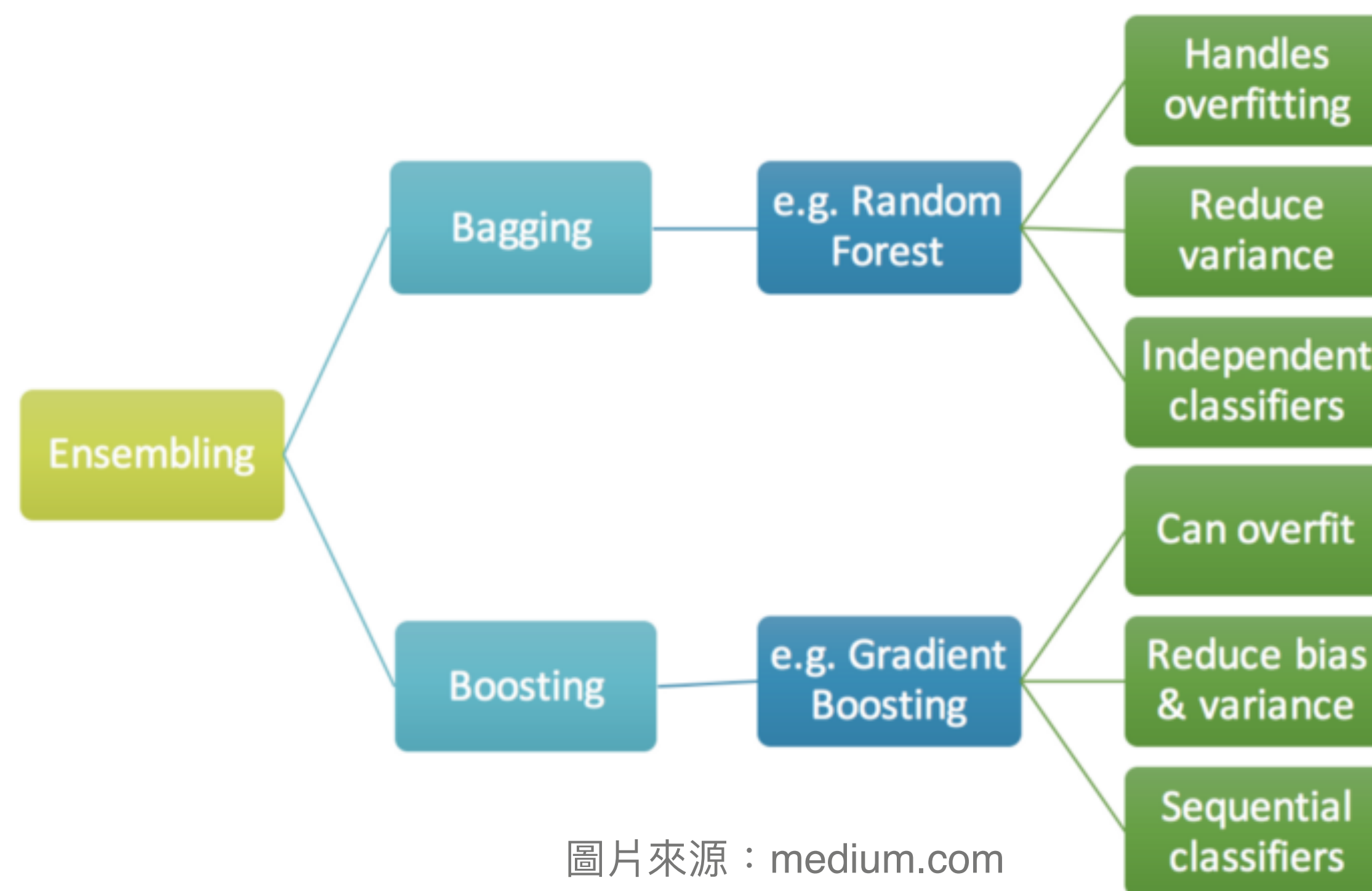
- 每次生成樹都是要修正前面樹預測的錯誤，並乘上 learning rate 讓後面的樹能有更多學習的空間



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

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



圖片來源：medium.com

充電時間 Brain Charge

請跳出PDF至官網Sample Code & 作業
進行今日作業

