

Day 43

機器學習

隨機森林



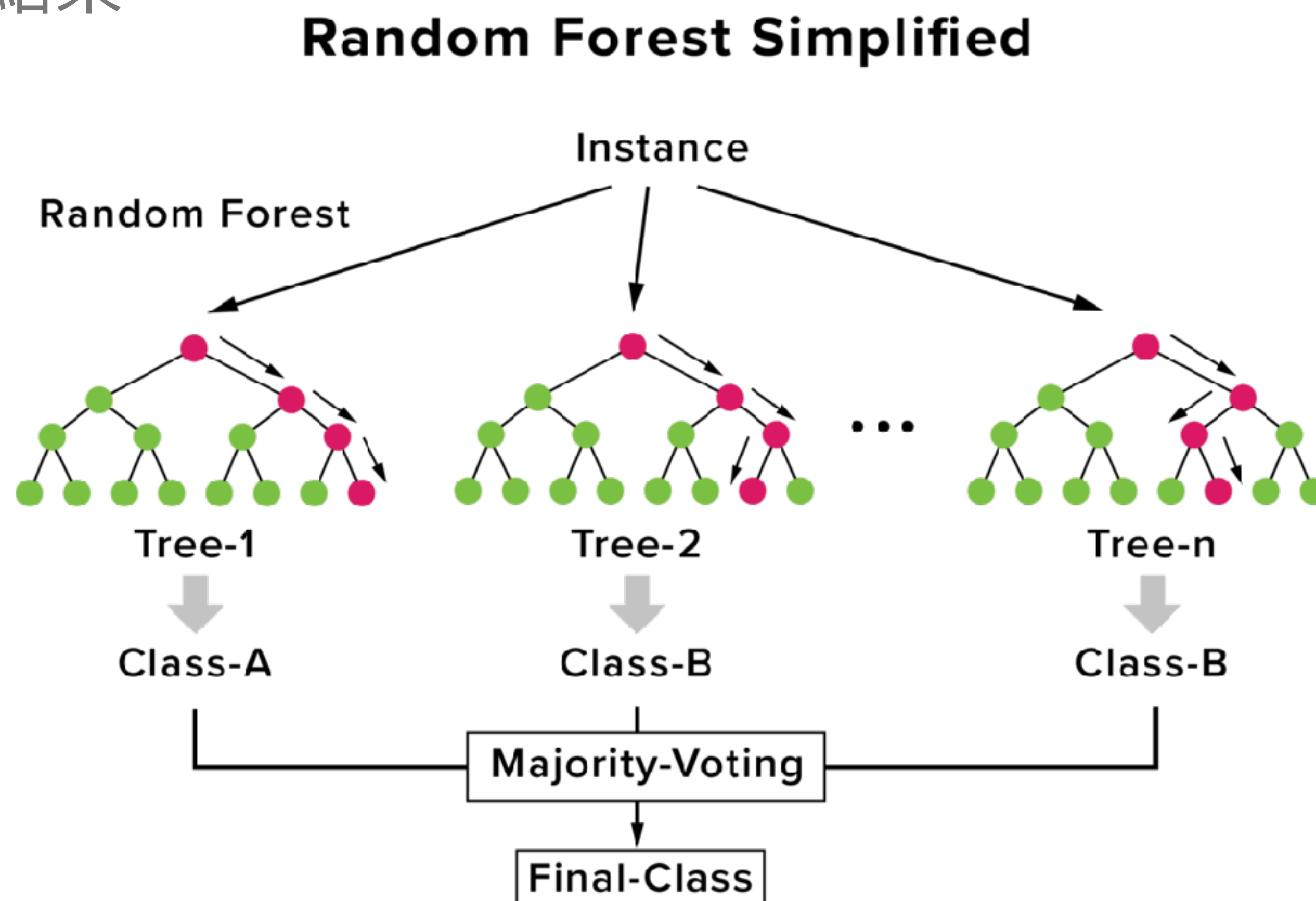
決策樹的缺點

- 若不對決策樹進行限制 (樹深度、葉子上至少要有多少樣本等)，決策樹非常容易 Over-fitting
- 為了解決決策樹的缺點，後續發展出了隨機森林的概念，以決策樹為基底延伸出的模型



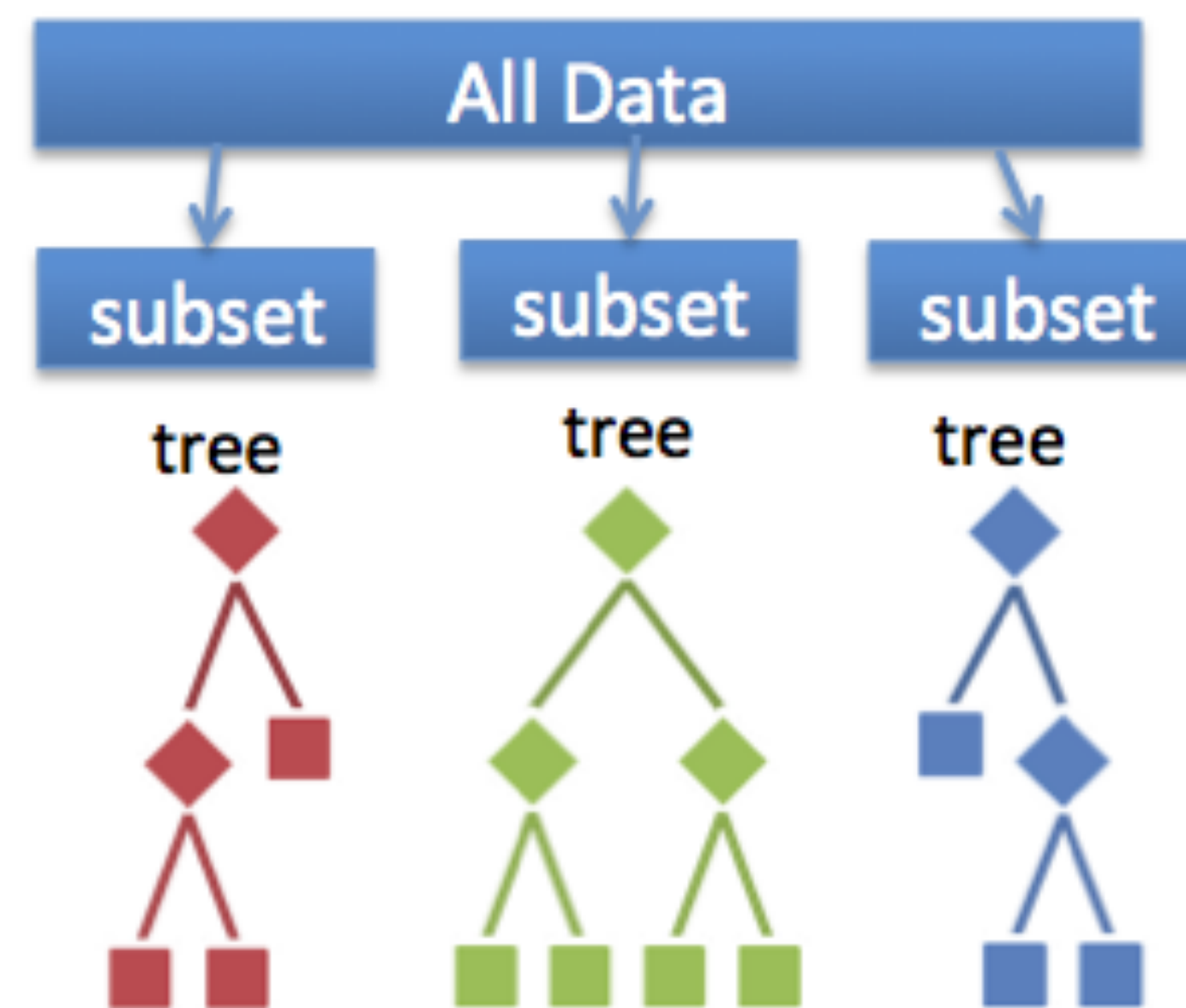
集成模型 - 隨機森林 (Random Forest)

集成 (Ensemble) 是將多個模型的結果組合在一起，透過投票或是加權的方式得到最終結果



隨機森林 (Random Forest), 隨機在哪？

- 決策樹生成時，是考慮所有資料與特徵來做切分的
- 而隨機森林的每一棵樹在生成過程中，都是隨機使用一部份的訓練資料與特徵
代表每棵樹都是用隨機的資料訓練而成的



解題時間 It's Your Turn

請跳出PDF至官網Sample Code & 作業
開始解題

