

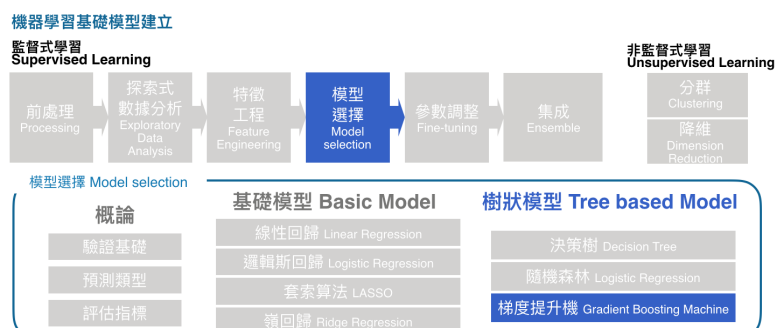
D45 : tree based model - 梯度提升機 (Gradient Boosting Machine) 介紹

[簡報閱讀](#)[範例與作業](#)[問題討論](#)[梯度提升機](#)[知識地圖](#)[本日知識點目標](#)[梯度提升機 \(Gradient Boosting Machine\)](#)[Gradient boosting 與 Random Forest 的差別？](#)[Bagging 與 Boosting 的差別](#)[延伸閱讀](#)[充電時間](#)

梯度提升機



知識地圖



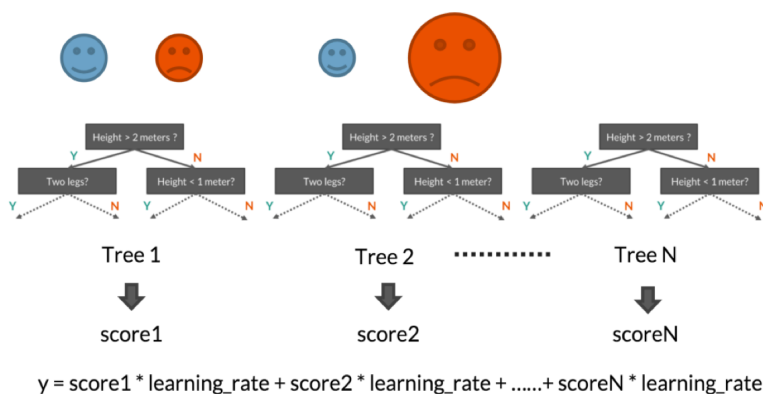
本日知識點目標

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!
- 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



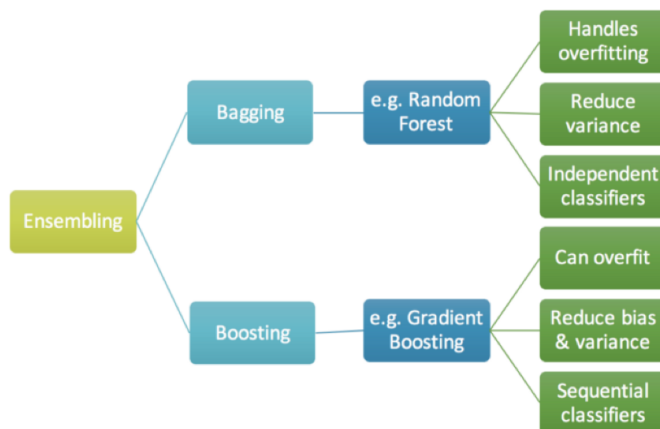
Gradient boosting 與 Random Forest 的差別？

Gradient boosting

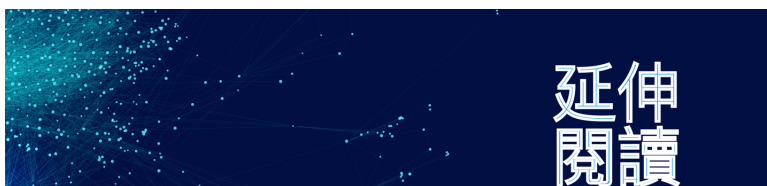
Gradient boosting 因為下一棵樹是為了修正前一棵樹的錯誤，因此每一棵樹皆有相關聯

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



延伸閱讀



- 梯度提升機原理 - 簡體
 - 連結：
<https://ifun01.com/84A3FW7.html>
 - 文章中的殘差就是前面提到的 Loss，從範例中了解殘差是如何被修正的

型是怎麼樣進行優化

Kaggle Winning Solution Xgboost Alg...



- XGBoost 數學原理 slides - 英文
 - 連結：
<https://homes.cs.washington.edu/~tqchen/pdf/BoostedTree.pdf>
 - 了解 XGBoost 的目標函數數學推導

充電時間



請跳出PDF至官網Sample Code & 作業
進行今日作業



[下一步：閱讀範例與完成作業](#)

