

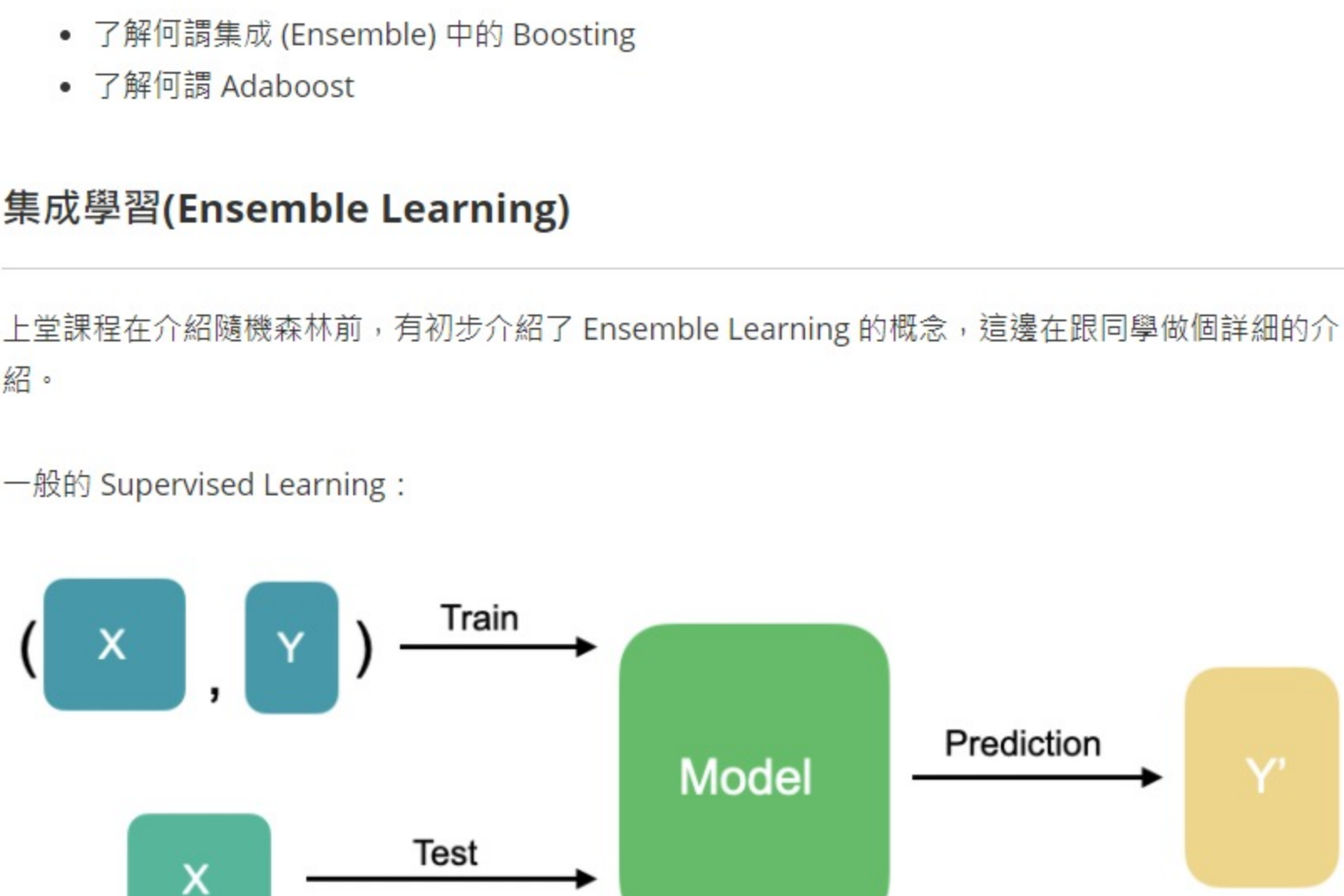
D26 : Adaboost 演算法



- 重要知識點 >
- 集成學習(Ensemble Learning) >
- Boosting >



重要知識點



- 了解何謂集成 (Ensemble) 中的 Boosting
- 了解何謂 Adaboost

集成學習(Ensemble Learning)

上堂課程在介紹隨機森林前，有初步介紹了 Ensemble Learning 的概念，這邊在跟同學做個詳細的介紹。

一般的 Supervised Learning：



- 當只有單一模型預測時，容易陷入 variance-bias trade-off 的問題中
- 單一模型陷入局部低點(local optimal)時，模型無法獲的最佳結果

Ensemble

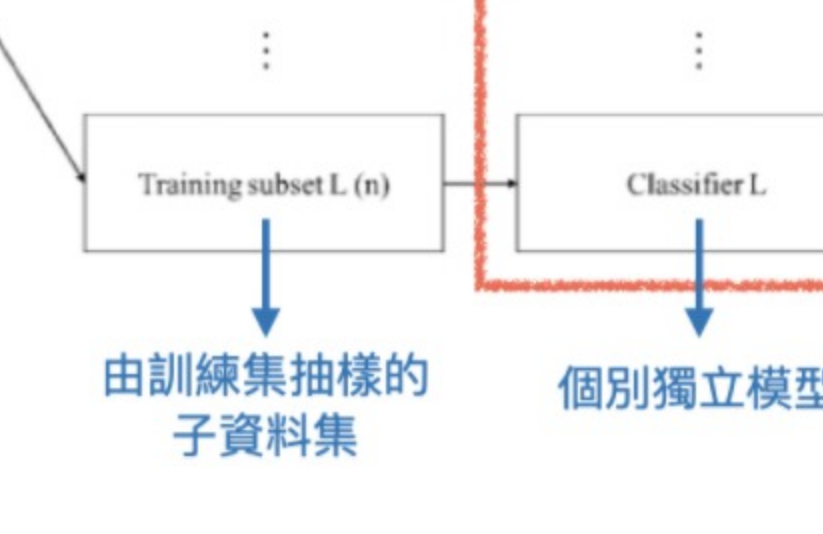
- 透過集結不同的模型，增加複雜度，減少 variance
- 透過平均或投票，集結眾多模型的預測，減少單一模型的重點

集成學習有個重要的論點在於，所有用來集成的模型必須夠歧異(Diverse)。

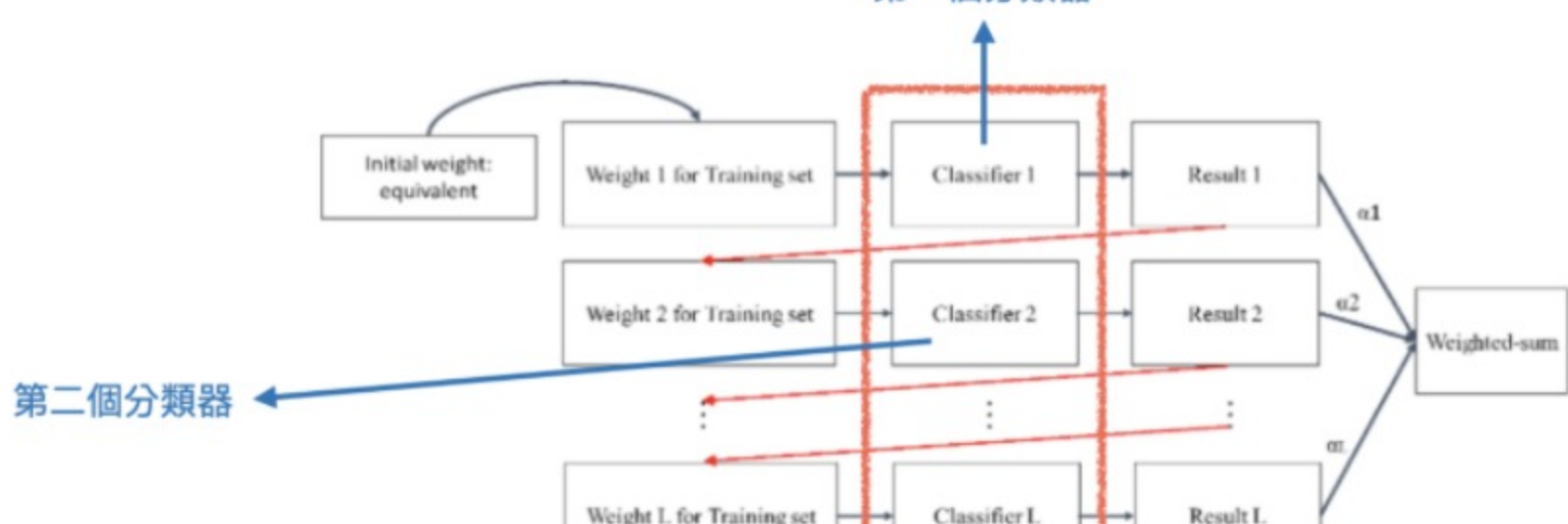
模型沒有歧異性，無法分出新的資料 (粉紅色方框)



模型有歧異性，集成後可分出新資料 (粉紅色方框)



Ensemble 中有不同的集成手法，最常見的有像是上個章節介紹過的 Bagging，也是 Random Forest 採用的集成方法。



而這章節主要要介紹的 Adaboost，則是另外一個常見集成手法 Boosting，Boosting 是一種 Bagging 的變形，其背後的思想在於“Boost 模型的表現”，將模型以序列的方式串接在一起，透過加強學習前一步的錯誤，來增強這一步模型的表現



Boosting

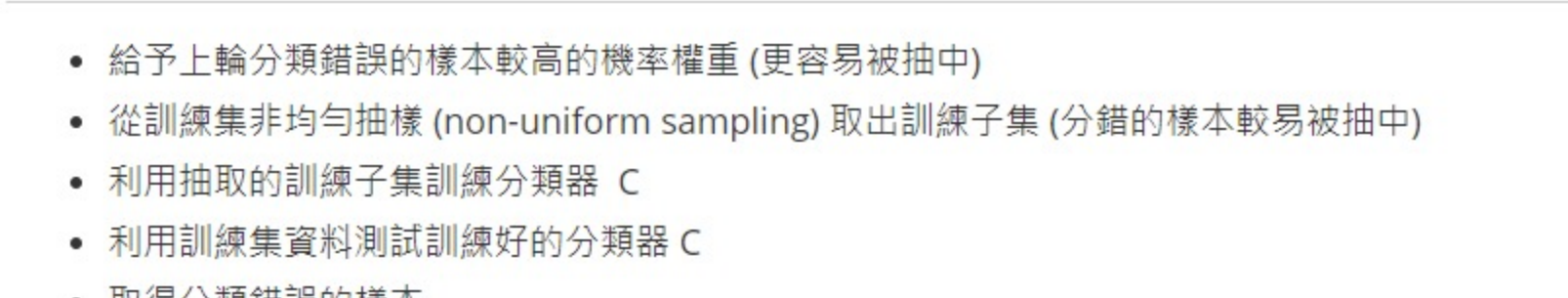
透過對訓練集進行取樣 (re-sample) 得到不同獨立的子訓練集 (subset)，利用這些子集對模型進行訓練，已取得個別不同的模型來進行 Ensemble，Boosting 也是使用這類的方式來進行訓練，與 Bagging 不同的是 Boosting 在取樣是會著重在**分錯**的樣本上。

接下來讓我們來看看 Boosting 演算法的步驟。

- 從訓練集均勻抽樣 (uniform sampling) 取出訓練子集
- 利用抽取的訓練子集訓練分類器 A
- 利用訓練集資料測試訓練好的分類器 A
- 取得分類錯誤的樣本



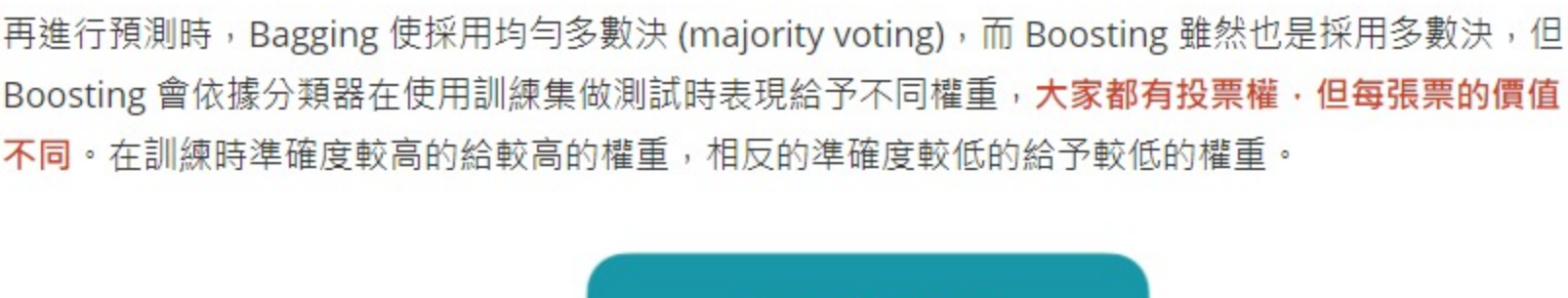
- 給予上輪分類錯誤的樣本較高的機率權重 (更容易被抽中)
- 從訓練集非均勻抽樣 (non-uniform sampling) 取出訓練子集 (分錯的樣本較易被抽中)
- 利用抽取的訓練子集訓練分類器 B
- 利用訓練集資料測試訓練好的分類器 B
- 取得分類錯誤的樣本



- 給予上輪分類錯誤的樣本較高的機率權重 (更容易被抽中)
- 從訓練集非均勻抽樣 (non-uniform sampling) 取出訓練子集 (分錯的樣本較易被抽中)
- 利用抽取的訓練子集訓練分類器 C
- 利用訓練集資料測試訓練好的分類器 C
- 取得分類錯誤的樣本



再進行預測時，Bagging 使用均勻多數決 (majority voting)，而 Boosting 雖然也是採用多數決，但 Boosting 會依據分類器在使用訓練集做測試時表現給予不同權重，**大家都投票權，但每張票的價值不同**。在訓練時準確度較高的給較高的權重，相反的準確度較低的給予較低的權重。



文字看不懂怎麼辦？讓我們以圖是來說明 Adaboost 的訓練流程



Adaboost 優點點

了解 Adaboost 的運作後，來看看 Adaboost 的特性與優點點

優點：

- 良好的利用弱分類器進行集成
- 具有高的精度
- 充分考慮不同分類器的權重分佈

缺點：

- 數據不平衡時會使分類精度下降
- 對資料的噪聲較為敏感
- 訓練所需的時間較長

Bagging V.S. Boosting

Bagging	Boosting
使用均勻採樣 (樣本被取得的機率相等)	非均勻採樣 (分類錯誤樣本被取得機率較高)
每輪訓練資料使用採樣的方式進行	每輪訓練資料使用權重分配的方式
模型為平行架構	模型為序列架構
使用均勻多數決	使用加權多數決

知識點回顧

在這章節我們學習到了

- 了解集成 (Ensemble) 中的 Boosting
- 了解 Adaboost 模型與其運作原理

延伸閱讀

網站：[Boosting介紹](#)

Bagging 介紹 YouTube 影片



網站：[Bagging與Boosting](#)

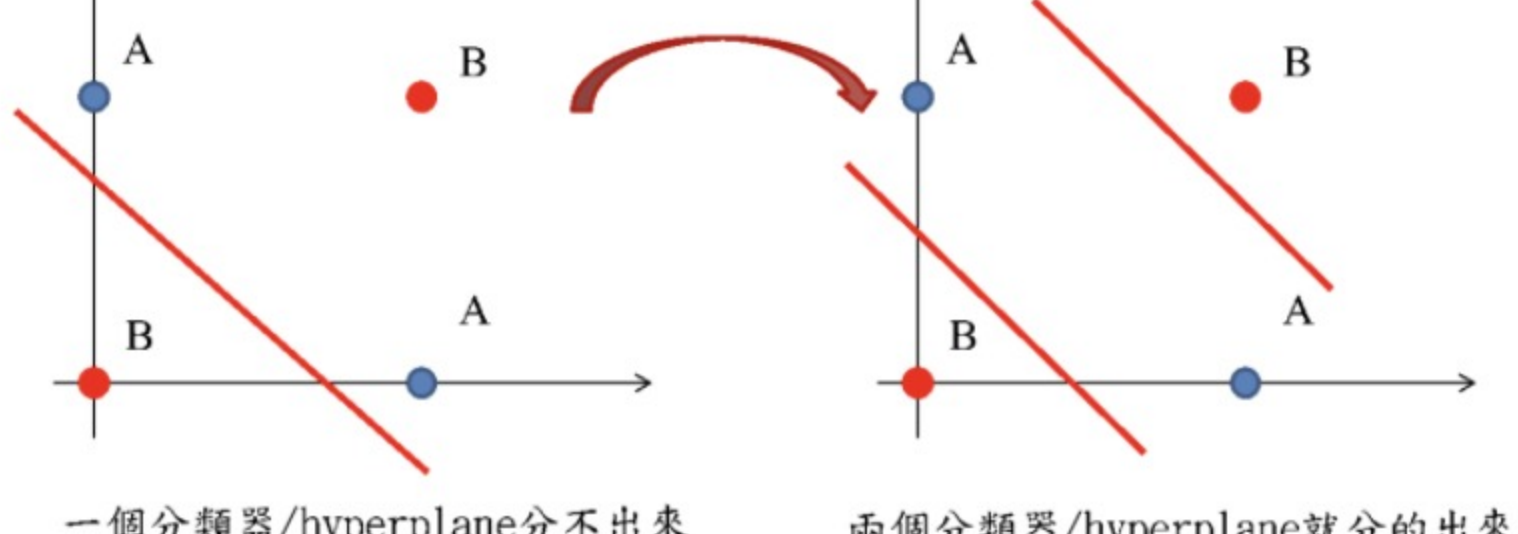
Bagging 與 Boosting 綜合介紹

機器學習: Ensemble learning之 Bagging、Boosting和AdaBoost

Tommy Huang Follow Jun 20, 2018 · 9 min read



Bagging、Boosting和AdaBoost (Adaptive Boosting)都是Ensemble learning(集成學習)的方法(手法)。Ensemble learning在我念書的時後我比較喜歡稱為多重辨識器，名稱很直覺，就是有很多個辨識器。其概念就是「三個臭皮匠勝過一個諸葛亮」，如果單個分類器表現的很好，那麼為什麼不用多個分類器呢？



Ensemble Learning基本條件是:每個分類器之間應該要有差異，每個分類器準確率需大於0.5。

如果用的分類器沒有差異，那只是用很多個一樣的分類器來分類，結果合成起來是沒有差異的。如果分類器的精度 $p < 0.5$ ，隨著ensemble規模的

[下一步：完成作業](#)