

② 對於缺失值處理

特徵值缺失下進行劃分特徵的選擇??

(即如何計算訊息增益??)

⇒ 有缺失值特徵, 用無缺失的樣本子集所占比
重新計算

選定該劃分特徵, 對於缺失該特徵值的樣本如何處理
(把此樣本劃分到哪一節葉中)

⇒ 將樣本同時劃分到所有子節葉中, 不違反
調整樣本的权重。

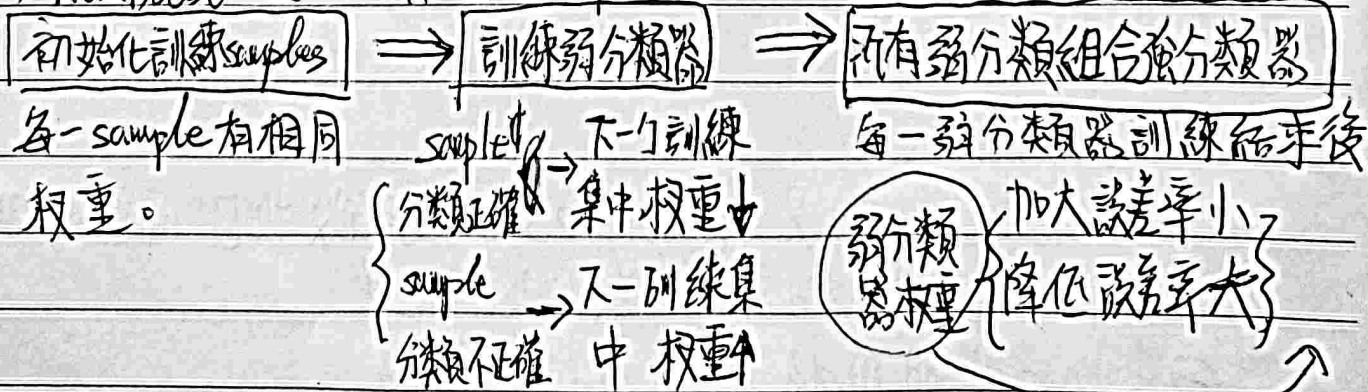
(以不同概率劃分到不同節葉中)

(三) Adaboost

(3.1) 何謂 Adaboost

為 boosting 算法的擴展, 增加「自適應」: 前一個基本分類器分錯
samples 會得到加強, 加權后的全体 samples 再次被用來訓練下一分
類器。同時, 在每一輪中加入一個新的弱分類器, 直到達到某一個
預定的足夠小的錯誤率, 或達到預先指定的最大迭代次數

(3.2) Adaboost 迭代算法 3 steps:



(3.3) 補充: 集成學習的偏差(Bias) & 方差(Variance)

Bias: 預測值與真實值之差。

Variance: 預測值作為隨機變量的離散程度。