

<https://www.analyticsvidhya.com/blog/2016/01/xgboost-algorithm-easy-steps/>

<http://leoyeh.me:8080/2017/11/19/%E8%B3%87%E6%96%99%E5%88%86%E6%9E%90-XGBoost-1/>

過適問題 (Overfitting)，所謂過適問題就是當模型在訓練資料集上的誤差低，但是在驗證資料集上的誤差非常高，最根本的辦法就是降低模型複雜度，像是增加訓練資料、減少訓練使用的特徵數量和提高正規化程度，因為我們對於好模型的判斷依據是準確且簡單為主，主要會以機器學習中的偏差 (Bias) 和變異 (Variance) 進行判斷模型是否有不足或過適的問題發生。在統計學中有個概念為偏差和變異之權衡 (Bias-Variance tradeoff)，簡單來說，偏差可以理解為若我們有無限資料時，訓練出最佳模型的誤差，我們會儘可能透過目標函數盡可能去擬合訓練資料，此時會有較少的偏差，而變異則是因為我們有限資料中隨機性所帶來的誤差，此時正規化則是讓模型盡可能簡單，因為當模型越簡單則有限資料擬合出來結果的隨機性較小，不易發生過適問題，使得最後模型的預測更加穩定。