Decision Tree Report

資工三 A_108502541_張凱博

_feature_split :

Design : 用 for 迴圈跑過每一筆特徵,並用每一筆特徵中的不重複資料當臨界點去計算 Information Gain 來找到當前最好的分割特徵及分割臨界值。

Goal : 找出當前最好的分割位置。

_build_tree :

Design : 建立一個 root_node 並用剛剛建好的_feature_split 來找分割點將資料分割，並遞迴將左右節點都跑_build_tree 藉此將整棵樹建立完整。

Goal : 將決策樹的模型建立起來。

_find_min_alpha :

Design : 運用 stack 將父節點的左右直點存入，一個點一個點的尋找，只要當前節點的 alpha 值較小，則更改儲存節點，藉此尋找最小 alpha 的最佳剪枝位置。

Goal : 遍歷整顆樹的節點找尋最小的 alpha,並回傳其位置當剪枝點。

_prune :

Design : 用運前面建好的_compute_alpha & _find_min_alpha 來找到最佳剪枝點，並將其節點的左右節點丟棄,將左右指向 None 藉此完成剪枝。

Goal : 進行剪枝的動作，來減少過度擬合。

The effect of different parameters

Prune_tree_times :

這個參數代表的是剪枝的次數，適度剪枝可以降低模型的過度擬合，但若過度剪枝會造成模型複雜度過低進而降低模型預測機率，而剪枝次數過低則會造成模型過度擬合增加訓練模型預測準越率，但測試準確率則會非常低。

max_depth:

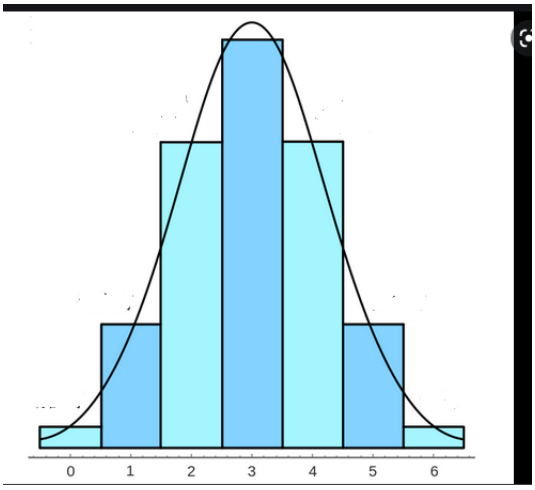
分割資料會將決策樹加深深度，這個參數是用來進行深度限制，若將 max_depth 設為很大很大，會造成嚴重的過度擬合，因為她會不斷的將資料分化到一個極度貼近

測試資料的決策樹，因次造成過度擬合。而 max_depth 設定為過小的話，會造成 underfitting，幾乎沒有將資料進行純化。

Result :

After prune tree, the testing accuracy become better….Why???

因為一開始的模型尚未進行剪枝，過度擬合，導致訓練預測率很高，但 test data 的準確率卻很低，因此進行剪枝後，將模型以最少增加錯誤的點進行修剪，不但不會讓錯誤率上升很多，還可以提高 test case 的預測準確率，因為模型不再那麼的過度擬合，可以更精準地進行預測，但過度的剪枝也會造成模型複雜度太低，test case 跟 training case 都不準確的狀況。



Decision tree before/after post pruning accuracy:

```
tree train accuracy: 0.966981
tree test accuracy: 0.670330
=============Cut=============
tree train accuracy: 0.962264
tree test accuracy: 0.703297
=============Cut=============
tree train accuracy: 0.962264
tree test accuracy: 0.703297
=============Cut=============
tree train accuracy: 0.957547
tree test accuracy: 0.714286
=============Cut=============
tree train accuracy: 0.952830
tree test accuracy: 0.725275
=============Cut=============
tree train accuracy: 0.938679
tree test accuracy: 0.736264
=============Cut=============
tree train accuracy: 0.924528
tree test accuracy: 0.736264
=============Cut=============
tree train accuracy: 0.924528
tree test accuracy: 0.736264
=============Cut=============
tree train accuracy: 0.900943
tree test accuracy: 0.747253
=============Cut=============
tree train accuracy: 0.886792
tree test accuracy: 0.747253
=============Cut=============
tree train accuracy: 0.886792
tree test accuracy: 0.747253
```