我把除了embarked外的資料都拿去train，在其他都沒改的情況下 test accuracy 變成 0.7430167597765364，就已經符合標準了，但是training accuracy 卻有 0.9859550561797753，代表 overfitting 了。我在網路上有找到一張圖表，他分析了training data中那些資料比較重要，我覺得除了原本的Sex, Age, Fare外，可以再加入Pclass。

一張含有 文字, 螢幕擷取畫面, 圖表, 字型 的圖片

自動產生的描述

決策樹的改進:

1. Criterion :用gini 或 entropy其實相差不大
2. Splitter: 我選擇用random比較可以防止overfitting
3. Max\_depth:我發現10大概是可以讓test accuracy比較高

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

模型的選擇:

我上網有査到有人把一些模型一起用，看哪個模型表現較好

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述





所以我就選擇用SVC 和 Random Forest

Randon Forest:用多個Decision Tree來分類，輸出的類別由個別樹的輸出眾數決定

1. Boostrap設為True，表示隨機抽樣

n\_estimators = 200, max\_depth = 9,這是我試出來的結果會最好

一張含有 文字, 字型, 螢幕擷取畫面, 印刷術 的圖片

自動產生的描述

SVC:只有linear的模型可以有最高的test accuracy

一張含有 文字, 字型, 螢幕擷取畫面, 印刷術 的圖片

自動產生的描述