HW1

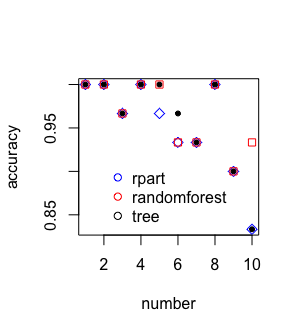
0513250 邱郁雯

1. 切割訓練資料與測試資料，比例為 80%：20%。

 切割iris資料庫，80%為訓練資料(iris.train)，20%為測試資料(iris.test)。

圖一

2. 分別利用tree套件（不修剪）、rpart套件（不修剪）、與randomForest進行10次測試，並比較其準確度（包含準確度的變化）。

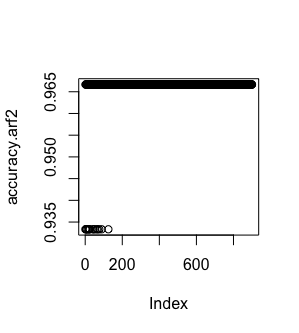
 將三個套件在迴圈進行10次測試後，tree、rpart、randomforest的準確度為圖二，並將測試次數與準確度畫出散佈圖（圖三），可看出測試次數越多，準確度沒有明顯提升，在一開始測試時準確度就已經很高，所以準確度變化不大。由圖四可知，tree平均準確度為0.96，rpart平均準確度為0.9533，randomforest平均準確度為0.9667，所以這三個套件的準確度randomforest > tree > rpart。

圖二 圖三

圖四

圖五

3. 利用randomForest套件，比較不同森林大小的準確度，並建議最合適的森林大小（分類樹的個數）。

使用for迴圈測試最佳的分類樹個數，從1顆到900顆測試，並且將每次測試之準確度畫成散佈圖，結果如圖七，1到900顆的分類樹準確度皆在90%以上，並且當分類數在200顆後，準確率高達100%。根據結果，我認為最佳分類樹的個數大於200都很適合。

圖六 圖七