

# Assignment 3 - Classification

Matthew Yu

2017/3

1. 利用 iris 資料庫，選擇期中 20% 為測試資料，其餘為訓練資料。

讀入資料、設定隨機種子並切割訓練與測試集

```
library(rpart)
library(e1071)
data(iris)
set.seed(100)
test.index <- sample(c(1:nrow(iris)),nrow(iris)*0.2)
iris.test <- iris[test.index,]
iris.train <- iris[-test.index]
```

2. 利用訓練資料，分別建立 Tree 以及 SVM 分類模式。

使用 rpart 套件中的 rpart 函數與 e1071 套件中的 svm 函數訓練及測試模型

```
iris_tree <- rpart(Species~., iris.train,control=rpart.control(cp=0))
tree.test <- predict(iris_tree, newdata=iris.test, type="class")
iris_svm <- svm(Species~., data=iris.train)
svm.test <- predict(iris_svm, newdata=iris.test[, -5])
```

3. 利用測試資料，分別測試 Tree 以及 SVM 模式的正確度。

以 Actual, Predicted 列聯表計算測試結果正確度

```
table(iris.test$Species, tree.test)

##           tree.test
##           setosa versicolor virginica
## setosa           8           0           0
## versicolor        0          11           1
## virginica         0           1           9

tree.accuracy.rate <- sum(diag(table(iris.test$Species, tree.test)))/nrow(iris.test)
table(iris.test$Species, svm.test)
```

```
##           svm.test
##           setosa versicolor virginica
## setosa           8           0           0
## versicolor        0          11           1
## virginica         0           0          10

svm.accuracy.rate <- sum(diag(table(iris.test$Species, svm.test)))/nrow(iris.test)
```

4. 輸出結果必須包含：Tree 的正確度、SVM 的正確度、哪個模式的正確度比較高。

```
Answer <- c(tree.accuracy.rate, svm.accuracy.rate)
names(Answer) <- c("Accuracy rate of Tree", "Accuracy rate of SVM")
Answer
```

```
## Accuracy rate of Tree  Accuracy rate of SVM
##           0.9333333      0.9666667
```

結果顯示在此一情況下 SVM 正確度較高