RandomForest

2022-04-28

library('stats19')

## Data provided under OGL v3.0. Cite the source and link to:  
## www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

library('dplyr')

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library('randomForest')

## randomForest 4.7-1

## Type rfNews() to see new features/changes/bug fixes.

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:dplyr':  
##   
## combine

mydata=iris  
str(mydata)

## 'data.frame': 150 obs. of 5 variables:  
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...  
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...  
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...  
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...  
## $ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...

index=sample(2,nrow(mydata), replace=TRUE,prob=c(0.7,0.3))  
training=mydata[index==1,]  
testing=mydata[index==2,]  
RFM <- randomForest(Species ~ .,data=training, importance=T, proximity=T)  
Species\_Pred=predict(RFM,testing)  
testing$Species\_Pred=Species\_Pred  
str(testing)

## 'data.frame': 50 obs. of 6 variables:  
## $ Sepal.Length: num 4.9 4.6 5.4 5 4.8 5.8 5.4 5.1 4.6 5.2 ...  
## $ Sepal.Width : num 3 3.1 3.9 3.4 3.4 4 3.4 3.7 3.6 3.4 ...  
## $ Petal.Length: num 1.4 1.5 1.7 1.5 1.6 1.2 1.7 1.5 1 1.4 ...  
## $ Petal.Width : num 0.2 0.2 0.4 0.2 0.2 0.2 0.2 0.4 0.2 0.2 ...  
## $ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Species\_Pred: Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...

CFM=table(testing$Species,testing$Species\_Pred)  
CFM

##   
## setosa versicolor virginica  
## setosa 21 0 0  
## versicolor 0 8 1  
## virginica 0 1 19