randomForest

Tao Huang

11/7/2021

Load the package

```
library("randomForest")
```

```
## randomForest 4.6-14
```

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

Load example data

```
data(iris)
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

Implement random forest algorithm

```
iris.rf <- randomForest(iris[,-5], iris[,5], prox=FALSE)
iris.rf
```

```
##
## Call:
##  randomForest(x = iris[, -5], y = iris[, 5], proximity = FALSE)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 2
##
##         OOB estimate of  error rate: 4.67%
## Confusion matrix:
##            setosa versicolor virginica class.error
## setosa         50          0         0        0.00
## versicolor      0         47         3        0.06
## virginica       0          4        46        0.08
```

A data proximity matrix is an important information source in random forests based data mining. The approach is based on measuring distance between two terminal nodes in a decision tree (Englund and Verikas, 2012).

```
iris.rf2 <- randomForest(iris[,-5], iris[,5], prox=TRUE)
iris.rf2
```

```
##
## Call:
##  randomForest(x = iris[, -5], y = iris[, 5], proximity = TRUE)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 2
##
##         OOB estimate of  error rate: 4%
## Confusion matrix:
##            setosa versicolor virginica class.error
## setosa         50          0         0        0.00
## versicolor      0         47         3        0.06
## virginica       0          3        47        0.06
```

REFERENCE

Englund, Cristofer, and Antanas Verikas. "A novel approach to estimate proximity in a random forest: An exploratory study." Expert systems with applications 39.17 (2012): 13046-13050.