JPL/Caltech Virtual Summer School in Big Data Analytics

Decision Trees and Random Forests

Exercise 1: Get Familiar with Random Forests (RF) in R

 # Install the Random Forest (RF) package if necessary
install.packages("randomForest")

 # load the RF package
library(randomForest)

 # take a look at the iris data
head(iris)

 # take a look at the class labesl
table(iris$Species)

 # Define the design matrix and the classifciation target
X <- iris[,1:4]
Y <- iris$Species

 # Learn a RF model
rf <- randomForest(x=X, y=Y, ntree=200, do.trace=T)

 # Print results
rf

 # Plot convergence of out-of-bag error
plot(rf)

 # investigate the vairable importance
importance(rf)
barplot(t(importance(rf)), col=4)