

Advanced Data mining

Lab : Random Forest

15 octobre 2019

1 Building and analysing random forests

1. Install the library *randomForest*.
2. Add to your data the variables *Title*, *AgeGroups*, *Family*, *FareGroup* and *Cabin2*.
3. Split your dataset into a training set and a test set.
4. Create a random forest using the default values of its parameters.
5. Evaluate the quality of your classifier using two measures.
6. What do you notice when you create and evaluate your RF several times ? How can we explain that ? How can we solve this "problem" ?
7. Extract a tree from the random forest and analyze its structure.
8. Try and compare many values of the RF parameters.
9. For the best choice of the parameters give the values of the variables' importance.

2 Extracting and analyzing rules

1. Install the library *inTrees*.
2. Use the functions of this library to extract and analyze the rules contained in your random forest.