

Advanced computational methods - Problem set 6

Hrvoje Stojic

March 2, 2016

1. Develop your own function for AdaBoost.M1 classifier discussed in the lectures. This is one of the most common algorithms and it would be good that you are more familiar with its inner workings.
 - Function should have the following name: **adaBoost**.
 - It should be stored in a file **adaBoost.R**.
 - Use a classification tree as a weak learner. You can use some existing function for this part of the code if you wish, e.g. **rpart** from the package with the same name.
 - Define it only for a binary classification problem.
 - Function should be able to deal with arbitrary number of dimensions.
 - You should have at least the following arguments to the function: **formula** (formula object for specifying the dependent variable and predictors), **data** (dataset that formula will operate on), **depth** (integer, max number of nodes in each tree), **noTrees** (number of iterations or trees). Please use exactly these names. Of course, you can add more arguments if you need to.
 - Verify that inputs are in a correct format (you can use **assertthat** package if you want).
 - Output should be a **named** list with at least this element - **predLabels** (vector of predicted labels for the observations in the dataset).
 - Finally, use SPAM dataset from the **Box/datasets** folder - generate training and test data, train your adaBoost on the training data and evaluate it on the test data. Do the same with adaBoost function from one of the packages and compare it with yours, e.g. **gbm** package. You should produce **adaBoost.pdf** that illustrates the evolution of training and test errors for your function and from the package for different number of iterations. Put this code in a file named **adaBoost_spam.R**.

Important notes

- Deadline is **March 11, at 12h**.

- Recall that you have to keep your code under version control in a repository at Github, I will pull the content of the repository at the designated time. Place all the files for this problem set in a folder in this repo under the name **PS6**.
- Create a simple text file with the name **Readme.md** in the problem set folder, if you want to leave me any instructions or messages regarding the problem set.
- Please, pay attention to the names of the files I instruct you to use. Names facilitate examining the code greatly, there is 30 of you and it takes me far more time if everybody produces different set of files with different names.
- I will give extra points for extra nice solutions!