

Machine Learning and Data Mining

Meta-learning, seminar

Maxim Borisyak

2016/09/22

National Research University
Higher School of Economics



HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

Using Titanic dataset (or any other vanilla one):

1. Using cross validation select optimal hyper-parameters for Random Forest and Extra Trees.
2. Plot train and test learning curves depending on number of trees:
 - for Random Forest;
 - for Extra Trees.
3. using bagging train a number of Decision Trees and stack them with Logistic Regression. Plot train and test learning curves for this stack depending on number of trees and their depth.
4. Calibrate Random Forest. Compare cross-entropy before calibration and after.