Ensembles – bagging

Comments on the illustrations below

- confidence intervals in the plots are based on 9-fold cross-validation and t-test with confidence level = 95%
- bagging is a random process, so another run would yield slightly different results
- experiment can be done using the demo code bagging.R
Ensembles – Random Forests

Comments on the illustrations below

- confidence intervals in the plots are based on 9-fold cross-validation and t-test with confidence level = 95%
- mtry value means the number of the randomly sampled features when a node of a decision tree is built
- both plots show that increasing the number of trees over 1,000 does not improve the classifier performance
- experiment can be done using the demo code rf.R
CRY -- Random Forest performance

avg accuracy (%) vs mtry parameter

different t values: 1000, 2000, 3000

ESSLLI '2015 -- Hladká & Holub
Illustrations to ensembles
Ensembles – AdaBoost

Comments on the illustrations below

- confidence intervals in the first plot are based on 9-fold cross-validation and t-test with confidence level = 95%
- the second plot focuses on the parameter boos of the boosting procedure; it shows the comparison between random runs with boos=T and another run with boos=F
  - boos=T means that a bootstrap sample of the training set is drawn in each AdaBoost iteration; thus, the learning process is randomized and each run gives different result
  - boos=F means that no bootstrapping is done and every training instance is used with its weights
- the third plot displays the same learning procedure that goes up to 500 trees in the ensemble
- experiment can be done using the demo code ab.R
CRY learning curve -- AdaBoost

avg accuracy (%) vs number of trees
CRY learning curve -- AdaBoost

four different runs with boos=T, and one run with boos=F