MA8701 Advanced methods in statistical inference and learning

L7: Random forest, Super Learner, Hyperparameter tuning

Mette Langaas IMF/NTNU

20 February, 2021

Ensembles - third act

So far

L5:

- Trees
- Many trees with bootstrap aggregation
- Many trees into a random forest (not finished)

L6: with Berent Å. S. Lunde (UiB)

- Boosting
- Xgboost
- Avoiding hyperparameter tuning in xgboost

Outline

- Ensembles
- Finishing the random forest from L5
- Super Learner
- Hyperparameter tuning

Left for L8: How to use statistical inference to give CI and compare predictions from different methods.

Ensembles - overview

(ELS Ch 16.1)

With ensembles we want to build *one prediction model* which combines the strength of *a collection of models*.

These models may be simple base models - or more elaborate models.

We have studied bagging - where we take a simple average of the prediction from many models (or majority vote), and the base models can be trees - or other type of models.

Random forest is a version of bagging, with trees made to be different (decorrelated).

We have studied boosting, where the models are trained on sequentially different data - from residuals or gradients of loss functions - and the ensemble members cast weighted votes. We have in particular looked into the xgboost variant of boosting, and also evaluated possible parameters to tune to optimize performance.

In L7 we also look into an ensemble built of elaborate models with the Super Learner and study some possible methods for tuning hyperparameters.