Bagging

Bagging (Bootstrap aggregating) is a *meta-learning technique* that involves creation of an ensemble of models based on random training sets and created from the original training set by sampling with replacement.

The final model is a simple average of the individual models within the ensemble.

In other words, bagging involves:

- replicating multiple **training sets** of equal sizes from the original training set. The size each set equals to the size of the original set
- definition of the respective **validation sets** as the samples not included in the training sets. By chance, about 37% of the samples will not included in a training set.
- **training multiple models** using a particular machine learning method (the same method and parameters for each model)
- for getting a validated prediction for a sample, calculate the prediction value as the average prediction by those models that had this compound in their validation sets

Bagging achieves two important goals: **validation** and **assessment of predictive uncertainty**, that is:

- Obtaining a correctly validated predictive statistics (similarly to cross-validation)
- Obtaining standard deviations for each prediction, which is possible because of using an ensemble of models rather that a single model.
  The standard deviation (referred to as BAGGING-STD) can be used, for calculation of prediction uncertainty and definition of applicability domain.