Bagging

Committees

When faced with a practical problem we are mostly interested in improving performance. We might want to do this by combining many methods.

For example, we might want to train M models and either take the average or let them vote.

This is called *a committee.*

Bootstrapping

- We want to train a set of classifiers and use the average result.
- We have to find a way to introduce *variability* between the models within the committee.
- We can use *bootstrap* datasets:
  - We sample N_m data points from $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ with replacement to obtain X_B.
  - Repeat M times to obtain M classifiers $y_m(\mathbf{x})$

For the theoretical foundation of this work see Dempter-Shafer Theory: https://en.wikipedia.org/wiki/Dempster%E2%80%93Shafer_theory

Aggregation

The committee's prediction is then given by

$$y_{COM}(\mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M} y_m(\mathbf{x})$$

This procedure is called bootstrap-and-aggregation or *bagging.*

Bagging

If the decisions by $y_m(\mathbf{x})$ are assumed to be independent then it can be shown that the error produced by $y_{COM}(\mathbf{x})$ is

$$E_{COM}(\mathbf{x}) = \frac{1}{M} E_{AV}$$

where E_{AV} is the average error of $y_m(\mathbf{x})$.

This, however, is not true as $y_m(\mathbf{x})$ are typically highly correlated. It can, however been shown that

$$E_{COM} \leq E_{AV}$$

Example of Bagging

The choice of weak learners lead to different examples of boosting, for example:

- Linear regressors – RANSAC (random sample consensus)
- Decision trees : Random forests