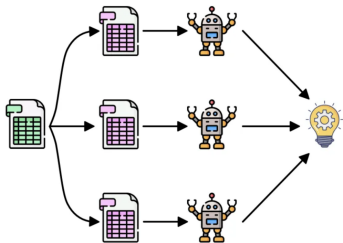


Random Forest

Bagging Ensembles

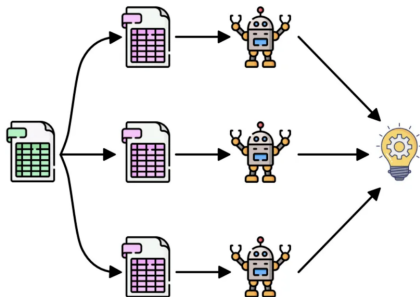


- Understand idea of bagging
- Be able to explain the connection between bagging and bootstrap
- Understand why bagging improves predictive performance



BAGGING

- Bagging is short for **B**ootstrap **A**ggregation
- **Ensemble method**, combines models into large “meta-model”; ensembles usually better than single **base learner**
- Homogeneous ensembles always use same BL class (e.g. CART), heterogeneous ensembles can use different classes
- Bagging is homogeneous



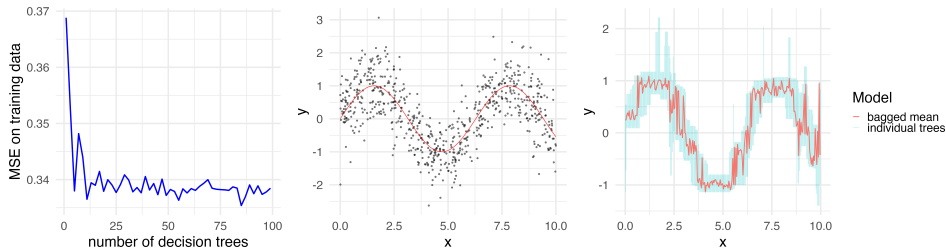
BAGGING PSEUDO CODE

Bagging algorithm: Training

- 1: **Input:** Obs. \mathbf{x} , trained BLs $\hat{b}^{[m]}$ (as scores $\hat{f}^{[m]}$, hard labels $\hat{h}^{[m]}$ or probs $\hat{\pi}^{[m]}$)
- 2: Aggregate/Average predictions

WHY/WHEN DOES BAGGING HELP?

- Bagging reduces the variability of predictions by averaging the outcomes from multiple BL models
- It is particularly effective when the errors of a BL are mainly due to (random) variability rather than systematic issues



- Increasing **nr. of BLs** improves performance, up to a point, optimal ensemble size depends on inducer and data distribution

MINI BENCHMARK

Bagged ensembles with 100 BLs each on spam:

Bagging seems especially helpful for less stable learners like CART

