

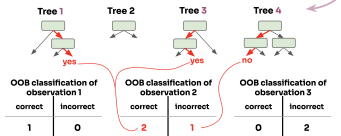
Introduction to Machine Learning

Random Forest

Out-of-Bag Error Estimate



ID	Color	Form	Length	Origin	Banana	OOB trees
1	yellow	oblong	14	imported	yes	{2}
2	brown	oblong	10	imported	yes	{1, 3, 4}
3	red	round	16	domestic	no	{2, 4}



Learning goals

- Understand the concept of out-of-bag and in-bag observations
- Learn how out-of-bag error provides an estimate of the generalization error during training

OUT-OF-BAG VS IN-BAG OBSERVATIONS



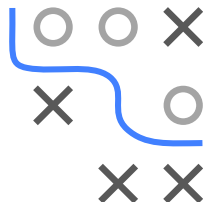
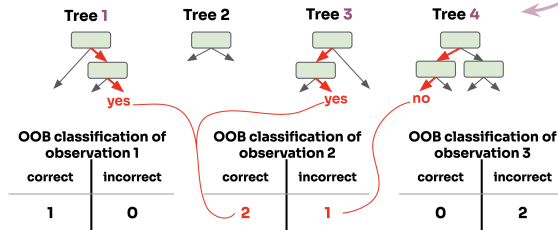
- IB observations for m -th bootstrap:
 $IB^{[m]} = \{i \in \{1, \dots, n\} \mid (\mathbf{x}^{(i)}, y^{(i)}) \in \mathcal{D}^{[m]}\}$
- OOB observations for m -th bootstrap:
 $OOB^{[m]} = \{i \in \{1, \dots, n\} \mid (\mathbf{x}^{(i)}, y^{(i)}) \notin \mathcal{D}^{[m]}\}$
- Nr. of trees where i -th observation is OOB:

$$s_{OOB}^{(i)} = \sum_{m=1}^M \mathbb{I}(i \in OOB^{[m]}).$$

OUT-OF-BAG ERROR ESTIMATE

Predict i -th observation with all trees $\hat{b}^{[m]}$ for which it is OOB:

ID	Color	Form	Length	Origin	Banana	OOB trees
1	yellow	oblong	14	imported	yes	{2}
2	brown	oblong	10	imported	yes	{1, 3, 4}
3	red	round	16	domestic	no	{2, 4}



OOB prediction $\hat{\pi}_{\text{OOB}}^{(2)} = 2/3$. Evaluating all OOB predictions with some loss function L or set-based metric ρ estimates the GE.

As we do not violate the **untouched test set principle**, $\widehat{\text{GE}}$ is not *optimistically* biased.

OUT-OF-BAG ERROR PSEUDO CODE

1: **Input:** $\text{OOB}^{[m]}, \hat{b}^{[m]} \forall m \in \{1, \dots, M\}$

3: Compute the ensemble OOB prediction for observation i , e.g., for regression:

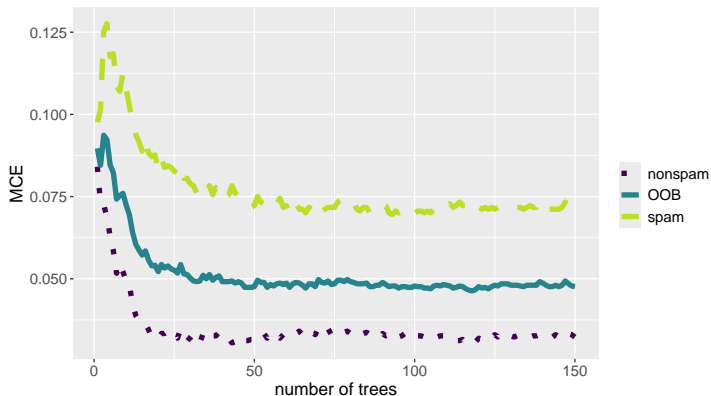
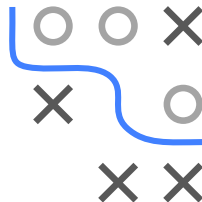
4: end for

$$\widehat{\text{GE}}_{\text{OOB}} = \frac{1}{n} \sum_{i=1}^n L(y^{(i)}, \hat{f}_{\text{OOB}}^{(i)})$$



USING THE OUT-OF-BAG ERROR ESTIMATE

- Gives us a (proper) estimator of GE, computable during training
- Can even compute this for all smaller ensemble sizes (after we fitted M models)



OOB ERROR: COMPARABILITY, BEST PRACTICE

OOB Size: The probability that an observation is out-of-bag (OOB) is:

$$\mathbb{P}\left(i \in \text{OOB}^{[m]}\right) = \left(1 - \frac{1}{n}\right)^n \xrightarrow{n \rightarrow \infty} \frac{1}{e} \approx 0.37$$

⇒ similar to holdout or 3-fold CV (1/3 validation, 2/3 training)

Comparability Issues:

- **OOB error** rather unique to RFs / bagging
- To compare models, we often still use CV, etc., to be consistent

Use the OOB Error for:

- Get first impression of RF performance
- Select ensemble size
- Efficiently evaluate different RF hyperparameter configurations

