# Introduction to Machine Learning

# Random Forest Out-of-Bag Error Estimate





#### 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**

ID	Color	Form	Length	Origin	Banana	ООВ
1	yellow	oblong	14	imported	yes	ІВ
2	brown	oblong	10	imported	yes	
3	red	round	16	domestic	no	predict
Bootstrapping to train tree 1						
ID	Color	Form	Length	Origin	Banana	Tree 1

imported

domestic

domestic

ves

no

no

14

16

16

× × ×

• IB observations for *m*-th bootstrap:  $IB^{[m]} = \{i \in \{1, ..., n\} | (\mathbf{x}^{(i)}, y^{(i)}) \in \mathcal{D}^{[m]}\}$ 

oblong

round

round

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

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

vellow

red

red

1

3

3

## **OUT-OF-BAG ERROR ESTIMATE**

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



× 0 0 × × ×

OOB prediction  $\hat{\pi}_{OOB}^{(2)} = 2/3$ . Evaluating all OOB predictions with some loss function *L* or set-based metric  $\rho$  estimates the GE. As we do not violate the **untouched test set principle**,  $\widehat{GE}$  is not *optimistically* biased.

### **OUT-OF-BAG ERROR PSEUDO CODE**

### Out-Of-Bag error estimation

- 1: Input:  $OOB^{[m]}, \hat{b}^{[m]} \forall m \in \{1, ..., M\}$
- 2: for  $i = 1 \rightarrow n \operatorname{do}$
- 3: Compute the ensemble OOB prediction for observation *i*, e.g., for regression:

$$\hat{t}_{\text{OOB}}^{(i)} = \frac{1}{S_{\text{OOB}}^{(i)}} \sum_{m=1}^{M} \mathbb{I}(i \in \text{OOB}^{[m]}) \cdot \hat{t}^{[m]}(\mathbf{x}^{(i)})$$

#### 4: end for

5: Average losses over all observations:

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



## 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)



× 0 0 × 0 × ×

## **OOB ERROR: COMPARABILITY, BEST PRACTICE**

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

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

 $\Rightarrow$  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

