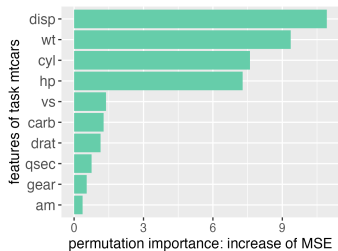
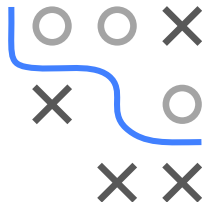


# Introduction to Machine Learning

## Random Forest Feature Importance



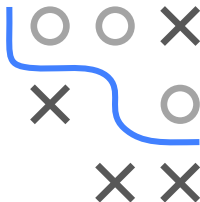
### Learning goals

- Understand that the goal of feature importance is to enhance interpretability of RF
- Understand FI based on feature permutation
- Understand FI based on improvement in splits

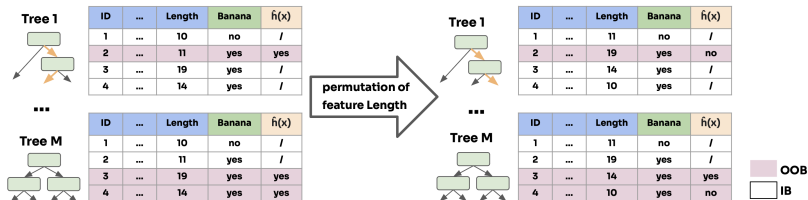
# PERMUTATION FEATURE IMPORTANCE

RFs improve accuracy by aggregating multiple decision trees but **lose interpretability** compared to a single tree. **Feature importance** mitigates this problem.

- How much does performance *decrease*, if feature is removed / rendered useless?
- We permute values of considered feature
- Removes association between feature and target, keeps marginal distribution
- Can obtain  $\widehat{GE}$  of RF (without and with permuted features) by predicting OOB data, to **efficiently compute FI during training**
- Avoids not only new models (if feature would be removed) but can already use “OOB test data” during training



ID	Color	Form	Origin	Length	Banana
1	yellow	round	domestic	10	no
2	brown	oblong	imported	11	yes
3	green	oblong	imported	19	yes
4	yellow	oblong	domestic	14	yes

[illegible]

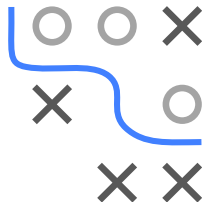
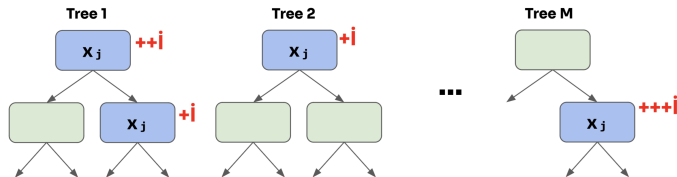
- ```

1: Calculate  $\widehat{\text{GE}}_{\text{OOB}}$  using set-based metric  $\rho$ 
2: for features  $x_j, j = 1 \rightarrow p$  do
3:   for Some statistical repetitions do
4:     Distort feature-target relation: permute  $x_j$  with  $\psi_j$ 
5:     Compute all  $n$  OOB-predictions for permuted feature data, obtain all  $\hat{f}_{\text{OOB}, \psi_j}^{(i)}$ 
6:     Arrange predictions in  $\hat{\mathbf{F}}_{\text{OOB}, \psi_j}$ ; Compute  $\widehat{\text{GE}}_{\text{OOB}, j} = \rho(\mathbf{y}, \hat{\mathbf{F}}_{\text{OOB}, \psi_j})$ 
7:     Estimate importance of  $j$ -th feature:  $\widehat{\text{FI}}_j = \widehat{\text{GE}}_{\text{OOB}, j} - \widehat{\text{GE}}_{\text{OOB}}$ 
8:   end for
9:   Average obtained  $\widehat{\text{FI}}_j$  values over reps
10: end for

```

# IMPURITY IMPORTANCE

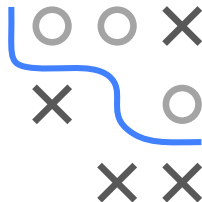
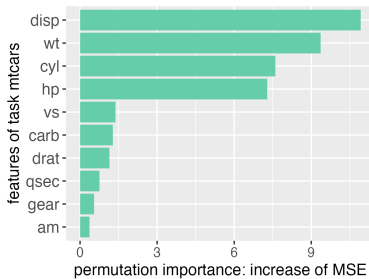
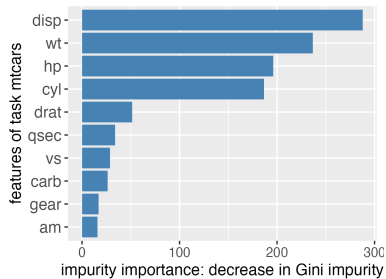
Alternative: Add up all *improvements* in splits where feature  $x_j$  is used.



- 
- 1: **for** features  $x_j, j = 1 \rightarrow p$  **do**
  - 2:     **for** all models  $\hat{b}^{[m]}, m = 1 \rightarrow M$  **do**
  - 3:         Find all splits in  $\hat{b}^{[m]}$  on  $x_j$
  - 4:         Extract improvement / risk reduction for these splits
  - 5:         Sum them up
  - 6:     **end for**
  - 7:     Add up improvements over all trees for FI of  $x_j$
  - 8: **end for**
-

# IN PRACTICE / OUTLOOK

Let's compare both FI variants on `mtcars`:



- Both methods are **biased toward features with more levels** (i.e., continuous or categoricals with many categories) ► Strobl et al. 2007
- More advanced versions exist
- PFI and FI have been generalized, see our lecture on IML!