Introduction to Machine Learning

Random Forests: Feature Importance

compstat-lmu.github.io/lecture_i2ml

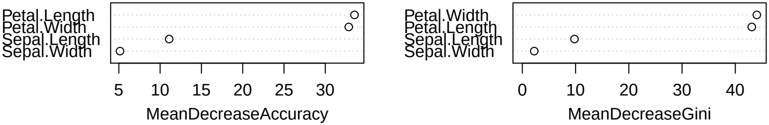
VARIABLE IMPORTANCE

- Single trees are highly interpretable
- Random Forests as ensembles of trees lose this feature
- Contributions of the different features to the model are difficult to evaluate
- Way out: variable importance measures
- Basic idea: by how much would performance of the random forest decrease if a specific feature were removed or rendered useless?

VARIABLE IMPORTANCE

Measure based on improvement in split criterion

for features x_j, $j = 1$ to p **do**
  for tree base learners $\hat{b}^{[m]}(x)$, $m = 1$ to M **do**
    Find all nodes $\mathcal{N}$ in $\hat{b}^{[m]}(x)$ that use x_j.
    Compute improvement in splitting criterion achieved by them.
    Add up these improvements.
  end for
  Add up improvements over all trees to get feature importance of x_j.
end for



VARIABLE IMPORTANCE

Measure based on permutations of OOB observations

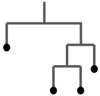
- While growing tree, pass down OOB observations and record predictive accuracy.
- Permute OOB observations of j-th feature. This destroys the association between the target and the permuted j-th feature.
- Pass down the permuted OOB observations and evaluate predictive accuracy again.
- The decrease of performance induced by permutation is averaged over all trees and is used as a measure for the importance of the j-th variable.

VARIABLE IMPORTANCE BASED ON PERMUTATIONS OF OOB OBSERVATIONS

	x_1	...	x_p	y	$\hat{y}$
1	1.4			1	1
2	2			0	0
3	1.55			1	1
4	1.72			0	0
5	1.89			1	0
...					
n	2.01			1	1

$$\boxed{\mathrm{R}_{\mathrm{m}}^{\mathrm{OOB}} = \frac{1}{|\mathrm{OOB}_{\mathrm{m}}|} \sum_{\mathrm{i} \in \mathrm{OOB}_{\mathrm{m}}} \mathrm{L}\left(\mathrm{y}^{(\mathrm{i})}, \hat{\mathrm{y}}_{\mathrm{m}}^{(i)}\right)}$$

Tree 1



Permuted OOB of tree 1

	x_1	...	x_p	y	$\hat{y}$
1	1.4			1	
2	2			0	
3	2.01			1	0
4	1.55			0	0
5	1.89			1	
...					
n	1.72			1	0

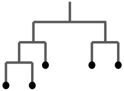
OOB obs. of x_1

	x_1	...	x_p	y	$\hat{y}$
1	1.4			1	
2	2			0	
3	1.55			1	1
4	1.72			0	0
5	1.89			1	
...					
n	2.01			1	1

$$\underbrace{R_1^{OOB} \text{ with permutation} - R_1^{OOB} \text{ without permutation}}_{\Delta R_1^{OOB}}$$

....

Tree M



Permuted OOB of tree 1

	x_1	...	x_p	y	$\hat{y}$
1	1.89			1	0
2	2			0	
3	1.4			1	0
4	1.72			0	
5	1.55			1	1
...					
n	2.01			1	

OOB obs. of x_1

	x_1	...	x_p	y	$\hat{y}$
1	1.4			1	1
2	2			0	
3	1.55			1	0
4	1.72			0	
5	1.89			1	1
...					
n	2.01			1	

$$\underbrace{R_M^{OOB} \text{ with permutation - } R_M^{OOB} \text{ without permutation}}_{\Delta R_{\mathrm{m}}^{OOB}}$$

$$\text{variable importance for } x_1 = \frac{1}{M} \sum_{m=1}^{M} \Delta R_m^{OOB}$$