Lecture 18

Ensemble Methods: Bagging

Murat Unal

Johnson Graduate School of Management
Cornell University

11/02/2021

Agenda

Bagging

Application in R

Ensemble methods

- We can overcome the weaknesses of a single tree by combining many individual trees.
- The following methods use trees as building blocks to construct more powerful prediction models.

1. Bagging
2. Random forests
3. Boosting

Bagging

- Individual decision trees are non-robust, i.e. have high variability.
- Averaging across observations reduces variance.
- Bagging(Bootstrap aggregation) uses this idea by taking repeated samples from the (single) training data set and averaging the results.
- This reduces the variance of the individual trees.

Bagging

- Bootstrapping involves repeatedly drawing independent samples from our data set (Z) to create bootstrap data sets $(Z_1, Z_2 \cdot Z_B)$.
- This sample is performed with replacement, which means that the same observation can be sampled more than once
- And each bootstrap sample will have the same number of observations as the original data set.

Bagging

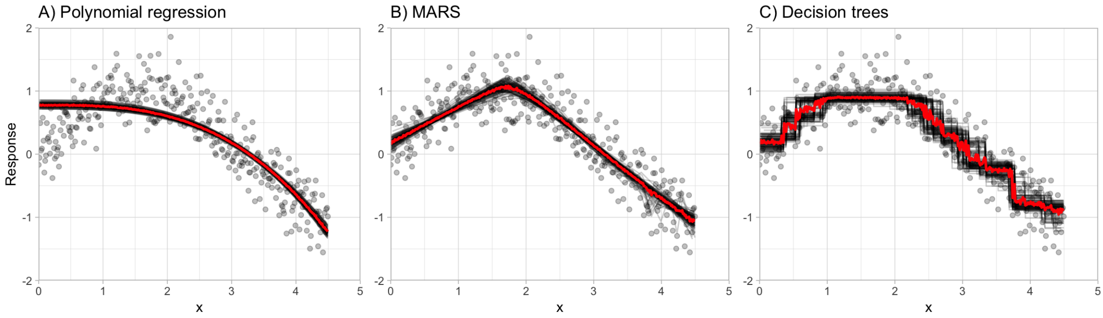
1. Create B different bootstrapped training data sets.
2. Train a decision tree $\hat{f}^b(x)$ on each of the samples.
3. Average all the predictions to obtain:

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}^b(x)$$

4. The final model is given by $\hat{f}_{bag}(x)$

Bagging

The effect of bagging 100 base learners. High variance models such as decision trees (C) benefit the most from the aggregation effect in bagging, whereas low variance models such as polynomial regression (A) show little improvement



Source: HOML

Bagging

- For regression trees: we typically apply bagging without pruning.
- We grow deep individual trees, which result in high variance and low bias.
- However, averaging ultimately reduces the variance.

Bagging

- For classification trees: for each test observation, we record the class predicted by each of the B trees, and take a **majority vote**, whereby the overall prediction is the most commonly occurring class among the B predictions.
- The number of trees is generally not critical with bagging. $B = 100$ has good performance.

Out-of-bag error estimation

- Bagging offers an easy way to obtain estimates of the test error without the need to do cross validation.
- For any bootstrapped sample, on average, each bagged tree makes use of around two-thirds of the observations.
- The remaining one-third of the observations not used to fit a given bagged tree are referred to as the out-of-bag (OOB) observations.

Out-of-bag error estimation

- We can predict the response for the ith observation using each of the trees in which that observation was OOB.
- This will yield around $B/3$ predictions for the ith observation, which we average.
- For each observation i:
  1. Find all samples S_i in which i was omitted from training
  2. Aggregate the $|S_i|$ predictions, using their mean or mode
  3. Compute the error, $y_i - \hat{f}_{i,OOB(x_i)}$
- This is a useful alternative for CV, because when B and n are large, CV will be computationally intensive.

Variable importance measure

- Bagging improves accuracy, but the price we pay is a loss in interpretability
- Now that we have an aggregate model, we no longer have an interpretable model.
- It is not clear anymore which variables are important in the model.

Variable importance measure

- We can obtain an overall summary of the importance of each feature using the RSS (for bagging regression trees) or the Gini index (for bagging classification trees).

Variable importance measure

- For bagged/RF regression trees, we record the total amount that the RSS is decreased due to splits over a given predictor, averaged over all B trees. A large value indicates an important predictor.
- Similarly, for bagged/RF classication trees, we add up the total amount that the Gini index is decreased by splits over a given predictor, averaged over all B trees.

References

Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2017)
An Introduction to Statistical Learning
Springer.
https://www.statlearning.com/

Ed Rubin (2020)
Economics 524 (424): Prediction and Machine-Learning in Econometrics
Univ, of Oregon.

Bradley Boehmke, Brandon Greenwell (2020)
Hands-On Machine Learning with R
Taylor & Francis Group
https://bradleyboehmke.github.io/HOML/