Bagging, Random Forests, Extra Trees

Extra Trees

Extra-Trees

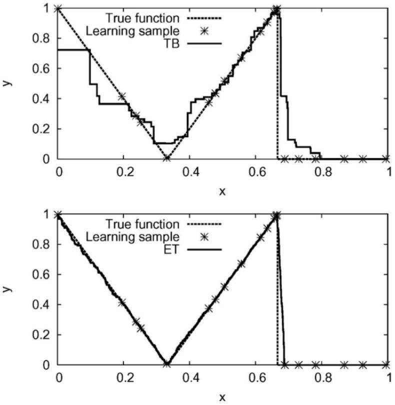
Taking randomization one step further...

1. Draw a random sample m from the p predictors (w/o Bootstrapping)
2. Draw a **random split** per feature
   - Numerical attribute a
     - Draw a cut-point a_c uniformly in $[a_{min}, a_{max}]$
     - Return the split $[a < a_c]$
   - Categorical attribute a
     - Compute $\mathcal{A}_s$, the subset of values of a that appear in the training data
     - Draw a subset $\mathcal{A}_1$ of $\mathcal{A}_s$ and a subset $\mathcal{A}_2$ of $\mathcal{A} \backslash \mathcal{A}_s$
     - Return the split $[a \in \mathcal{A}_1 \cup \mathcal{A}_2]$
3. Split node using the best of these random splits

$\rightarrow$ Extremely Randomized Trees (Geurts et al. 2006)

Extra-Trees

Figure: Comparison of Tree Bagging (TB) and Extra Trees (ET) models (example, Geurts et al. 2006)



Advantages of Extra-Trees

- Strong variance reduction
  - Random cut-points enforce de-correlation
  - *Totally randomized trees* with $\texttt{mtry} = 1$
- Small(er) bias?
  - Usage of full training sample
- Computational benefits
  - Simple node splitting procedure

Summary

- Resample-and-combine technique
- Bagging mimics averaging over several training sets
- Resampling utilized in various ML contexts
- Stabilizes predictions from high-variance methods (e.g., CART)
- Lower-level effect: Equalizes influence of outliers (Grandvalet 2004)

Software Resources

Resources for R

- Standard package to grow RFs: `randomForest`
- Fast implementation of RFs: `ranger`
- Extremely Randomized Trees: `extraTrees`
- Ensembles of Conditional Inference Trees: `cforest`
- Classification and Regression Training: `caret`
    - `https://topepo.github.io/caret/`

References

Berk, R. A. (2006). An Introduction to Ensemble Methods for Data Analysis. *Sociological Methods & Research, 34*(3), 263–295.

Biau, G., Scornet, E. (2015). *A Random Forest Guided Tour*. arXiv: 1511.05741.

Breiman, L. (2001). Random forests. *Machine Learning 45*(1), 5–32.

Breiman, L. (1996). Bagging Predictors. *Machine Learning, 24*(2), 123–140.

Geurts, P., Ernst, D., Wehenkel, L. (2006). Extremely Randomized Trees. *Machine Learning 63*(1), 3–42.

Grandvalet, Y. (2004). Bagging Equalizes Influence. *Machine Learning, 55*(3), 251–270.

Hall, P., Samworth, R. J. (2005). Properties of bagged nearest neighbour classifiers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67: 363–379.

Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer.

James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An Introduction to Statistical Learning*. New York, NY: Springer.

Strobl, C., Malley, J., Tutz, G. (2009). An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. *Psychological Methods 14*(4). 323–348.