Adaptive Bagging

Statistical Machine Learning

Veronica Burt

Bagging Recap

- Bagging takes a low bias, high variance estimator and reduces the variance.

- This is done by averaging together several estimators from bootstrap draws.

- If the estimators are highly correlated, we can only reduce the variance so much. Recall that the average of B random variables with correlation ρ has variance $\rho\sigma^2 + \frac{(1-\rho)\sigma^2}{B}$

Adaptive Bagging

- Adaptive bagging takes an estimator and reduces both the bias and the variance, provided that the estimator meets a certain condition.

- This is done by bagging, altering the output values using the out of bag samples, and bagging again using the altered output values. We can repeat this process as necessary.

- This can save on computation time; instead of growing a full unpruned tree for each bootstrap draw, we can grow smaller trees and have similar results.

Recap of the Bagging Process

To bag trees:

- Choose a large number for B, the number of bootstrap draws.

- Grow an unpruned tree on the bth bootstrap draw.

- Average the trees together.

Adaptive Bagging Process

To adaptively bag, there are multiple stages of bagging:

- Bag a fixed number B of trees (they do not have to be fully grown) using the same input values but with altered output values. Let the altered output values for the jth stage be denoted by $\{y_n^{(j)}\}$.

- Let $\widehat{y}_{n,b}$ denote the predicted values give by the bth tree grown. After the jth stage, $y_n^{(j+1)} = y_n^{(j)} - \bar{\widehat{y}}_{n,b}$, where $\bar{\widehat{y}}_{n.b}$ represents the average of the predicted values over the out of bag samples.

- If $\boldsymbol{x}$ is an input value not in the initial training set, then $f_{j,b}(\boldsymbol{x})$ denotes the prediction for $\boldsymbol{x}$ by the bth predictor at the jth stage. Then

  $f^{(j+1)}(\boldsymbol{x}) = f^{(j)}(\boldsymbol{x}) + \bar{f_{j,b}}(\boldsymbol{x})$, where $\bar{f_{j,b}}(\boldsymbol{x})$ represents the average of prediction for $\boldsymbol{x}$ over the first j stages.

- Repeat this process until the mean sum of squares of the new y's is greater than 1.1 times the minimum of the mean sum of squares of the y's in any of the previous stages. (The mean sum of squares of the y's is a measure of the residual bias).

- Use the predictor given by the end of the stage having the minimum mean sum of squares.

Weak Learning Condition

- Adaptive Bagging works only if the estimator fulfills a condition similar to the concept of a weak learner in classification.

- Basically, if the correlation between the predictor, f, and the operator (averaging over functions), is greater than $.5(1 + \epsilon)$, the condition is satisfied.

Empirical Results

Table 1 Bias and Variance--Unpruned CART

Data Set	Bias	Variance	Noise
Peak20	10.5	33.5	0.0
Friedman#1	3.4	10.7	1.0
Friedman#2+3	1.0	26.7	16.0
Friedman#3 -3	8.0	33.8	11.1

Empirical Results

Table 2 Bias-Variance for Bagging and Adaptive Bagging

Data Set	Bagging		Adaptive	
	Bias	Variance	Bias	Variance
Peak20	10.7	2.2	1.1	2.7
Friedman#1	3.8	1.4	1.2	1.9
Friedman#2	0.7	4.6	- - -	- - -
Freidman#3	7.6	5.9	6.2	7.4

Empirical Results

Table 10 Misclassification Errors (%)

Data Set	Two-Err	Min-Err	UP-Err	Ada-Err
diabetes	24,1	23.4 (3)	24.6	26.6
breast	5.6	3.9 (5)	4.2	3.2
ionosphere	7.0	6.6 (8)	7.7	6.4
sonar	23.0	14.1 (8)	14.9	15.6
heart (Clevld)	15.6	15.6 (2)	18.8	20.7
german credit	25.3	23.6 (7)	24.8	23.5
votes	4.5	3.7 (10)	4.6	5.4
liver	29.6	25.9 (6)	30.4	28.7

Advantages of Adaptive Bagging

- When Adaptive Bagging does not reduce bias, the process stops after one stage, so it is no less effective than bagging.

- It is most effective when the bias of the predictor is larger than the variance.

- It can save computation time; a tree with just one split takes $1/\log_2(N)$ as much computation time as growing the full unpruned tree.

- It tends to work well for classification problems.

Disadvantage of Adaptive Bagging

- Because there is dependence between estimates and the predictors selected, out of bag error estimates are biased for adaptive bagging.

- However, adaptive bagging still reduces bias of the predictor.

References

Brieman, Leo

Using Adaptive Bagging to Debias Regressions

Technical Report 546, February 1999.