

How Boosting the Margin Can Also Boost Classifier Complexity

Lev Reyzin and Robert E. Schapire

Boosting

Combine many “moderately inaccurate” base classifiers into a combined predictor

Generate a new “base classifier” in each round, constantly focusing on the hardest examples

The final predictor is the weighted vote of the base classifiers

Boosting tends not to overfit, even after many rounds

AdaBoost sets voting weights of each new generated base classifier to greedily reduce training error.

The Margins Theory

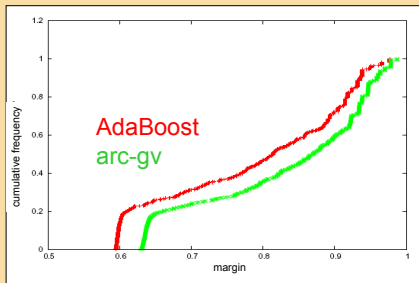
The margin of a hypothesis on a training example:
Its magnitude represents the confidence of the vote
It's positive if the vote gives the correct classification.
Otherwise it's negative.

the generalization error is at most:

$$\hat{\Pr}[\text{margin}_f(x, y) \leq \theta] + \tilde{O}\left(\sqrt{\frac{d}{m\theta^2}}\right)$$

Annotations:
 - $\tilde{O}\left(\sqrt{\frac{d}{m\theta^2}}\right)$: the VC dimension of the base classifier
 - m : number of training examples
 - θ : for any value of theta

Breiman's arc-gv: A Strong Case Against the Margins Theory

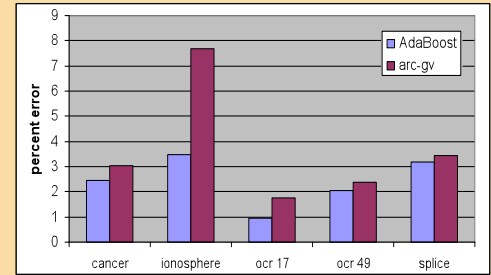


500 rounds of boosting on the “breast cancer” dataset using pruned CART trees as weak learners.

arc-gv greedily maximizes the minimum margin.

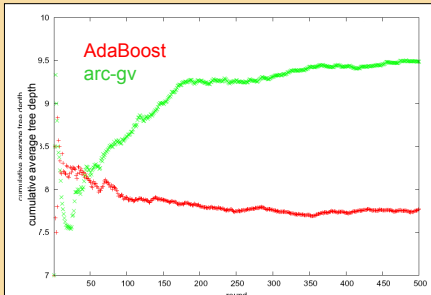
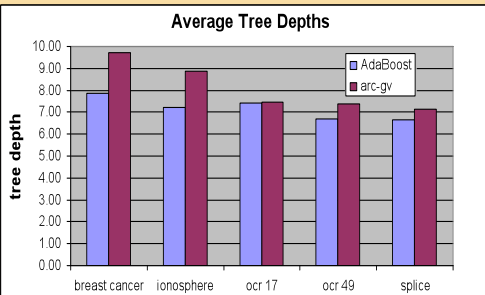
arc-gv's minimum margin converges to the optimal

Since arc-gv has uniformly higher margins and higher error than AdaBoost, Breiman concluded this refutes the margins theory



Test errors of AdaBoost and arc-gv with pruned CART trees as base classifiers

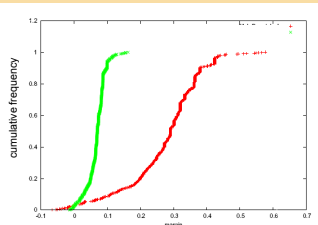
The Complexity Term in CART Trees: Supporting the Margin



differences of test and training errors per generated tree averaged over all CART trees in 500 rounds of boosting

	AdaBoost	arc-gv
cancer	3.5	4.1
ionosphere	8.9	9.9
ocr 17	1.9	2.0
ocr 49	3.7	4.1
splice	4.3	4.7

Decision Stumps Also Support the Margins Theory



	test error	
	arc-gv	AdaBoost
cancer	4.15	4.29
ionosphere	10.27	9.58
ocr 17	1.12	1.10
ocr 49	6.38	6.28
splice	7.22	6.79

Conclusions

Breiman's results do not actually contradict the margins theory.

Margins are important, but not always at the expense of other factors.

Our work adds one more piece to the margins puzzle.