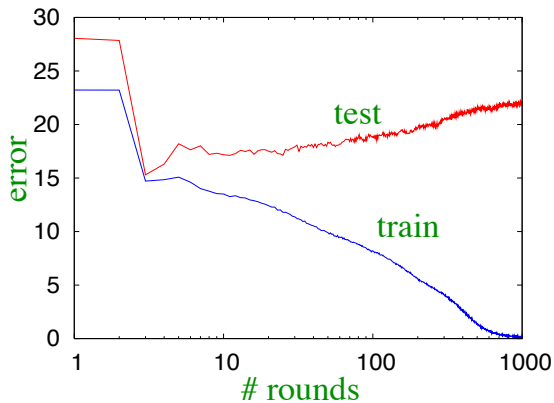
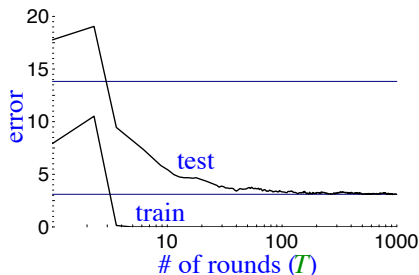


## Overfitting Can Happen



(boosting “stumps” on heart-disease dataset)

## Actual Typical Run



(boosting C4.5 on  
“letter” dataset)

- test error does **not** increase, even after 1000 rounds
  - (total size > 2,000,000 nodes)
- test error continues to drop even after training error is zero!

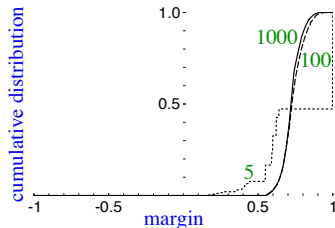
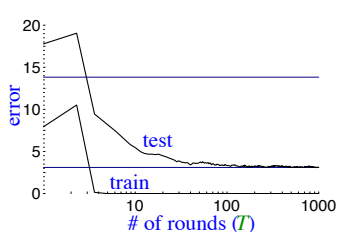
	# rounds		
	5	100	1000
train error	0.0	0.0	0.0
test error	8.4	3.3	3.1

- Occam’s razor **wrongly** predicts “simpler” rule is better

## Empirical Evidence: The Margin Distribution

- margin distribution

= cumulative distribution of margins of training examples



	# rounds		
	5	100	1000
train error	0.0	0.0	0.0
test error	8.4	3.3	3.1
% margins $\leq 0.5$	7.7	0.0	0.0
minimum margin	0.14	0.52	0.55