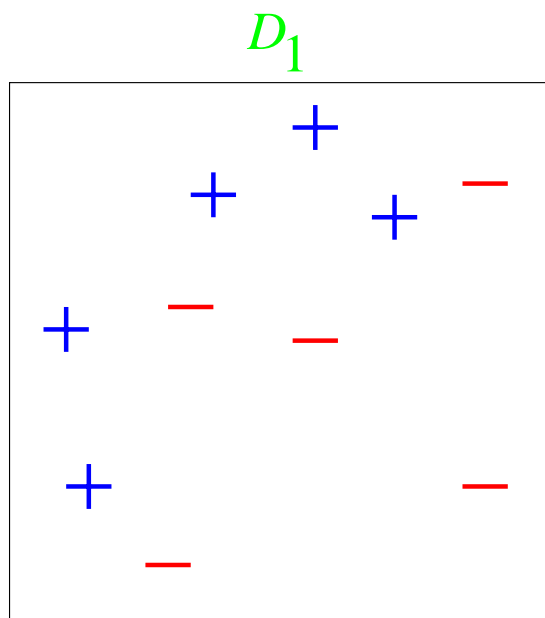
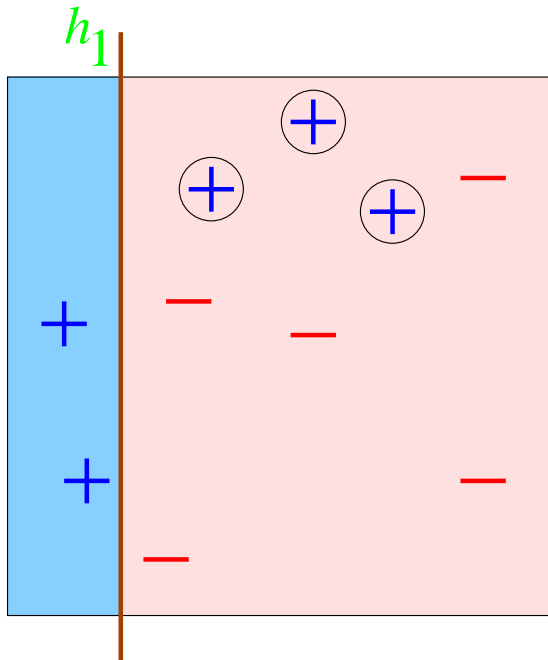


Toy Example



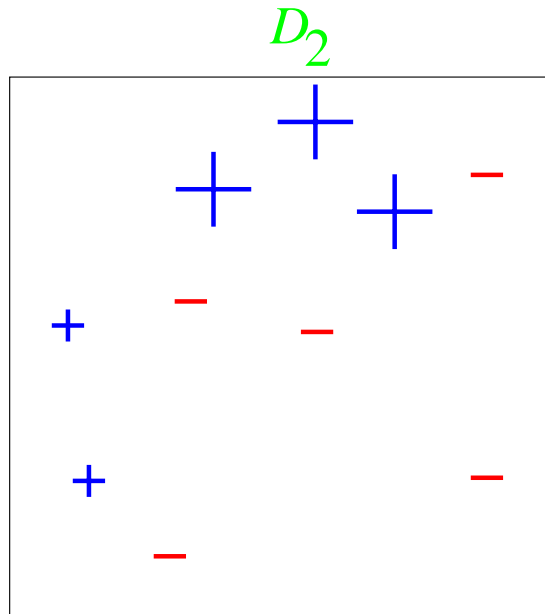
weak classifiers = vertical or horizontal half-planes

Round 1

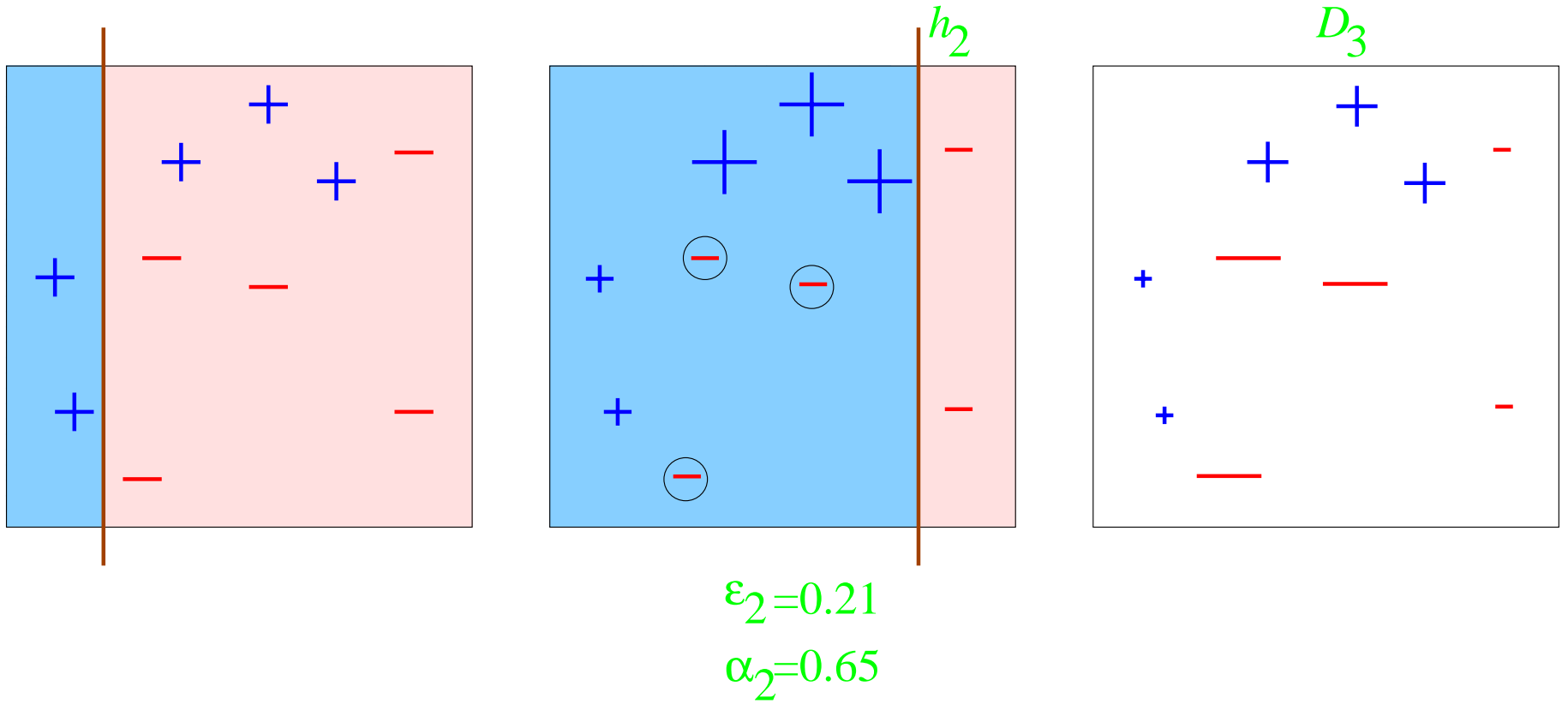


$$\varepsilon_1 = 0.30$$

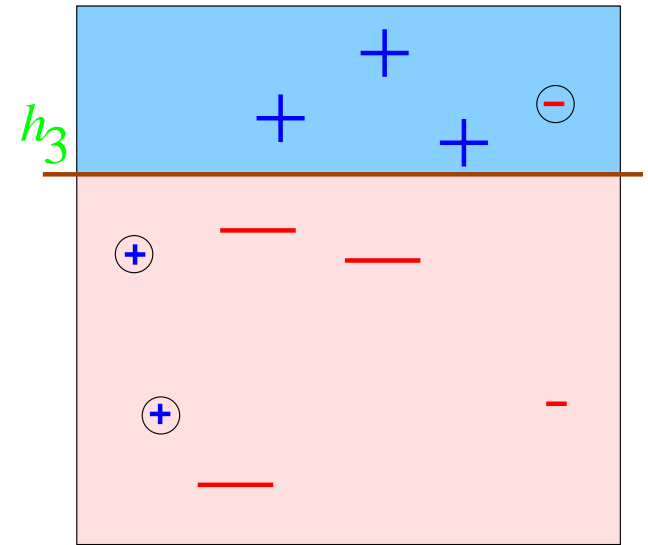
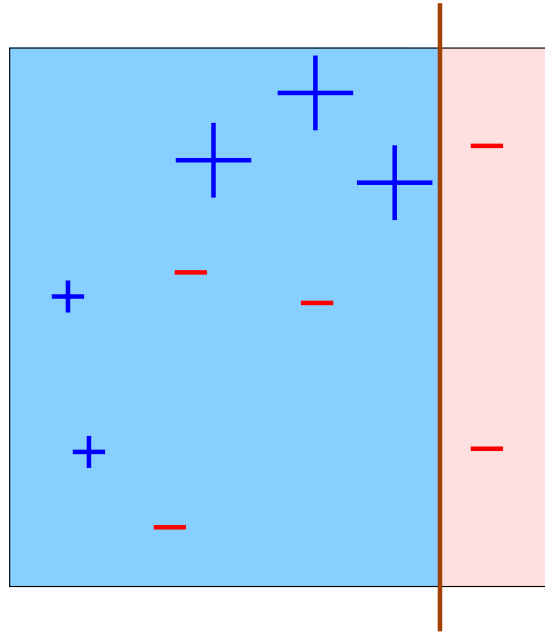
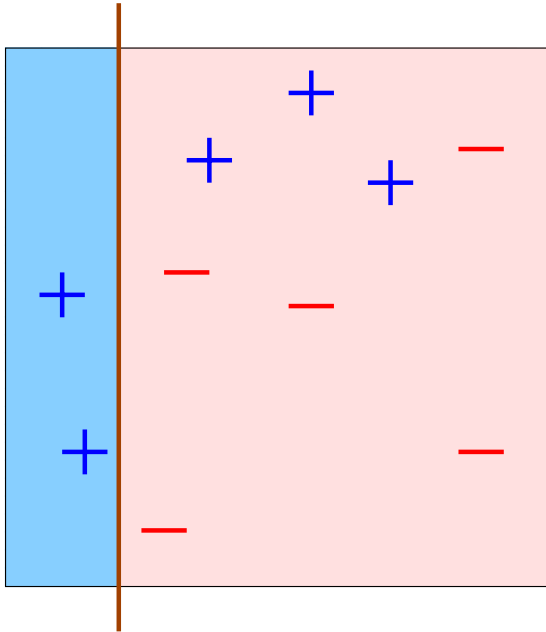
$$\alpha_1 = 0.42$$



Round 2



Round 3

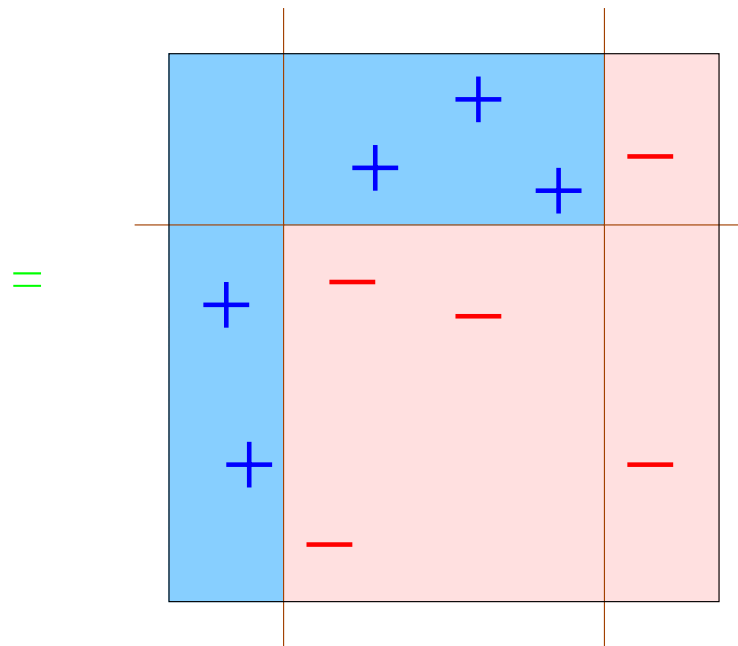


$$\epsilon_3 = 0.14$$

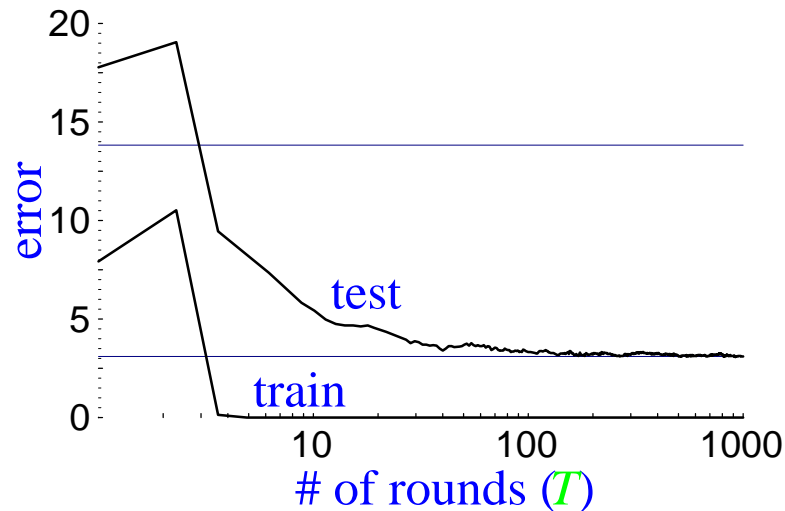
$$\alpha_3 = 0.92$$

Final Classifier

$$H_{\text{final}} = \text{sign} \left(0.42 \left(\begin{array}{|c|} \hline \text{blue} \\ \hline \end{array} \right) + 0.65 \left(\begin{array}{|c|} \hline \text{blue} \\ \hline \end{array} \right) + 0.92 \left(\begin{array}{|c|} \hline \text{blue} \\ \hline \end{array} \right) \right)$$



Actual Typical Run



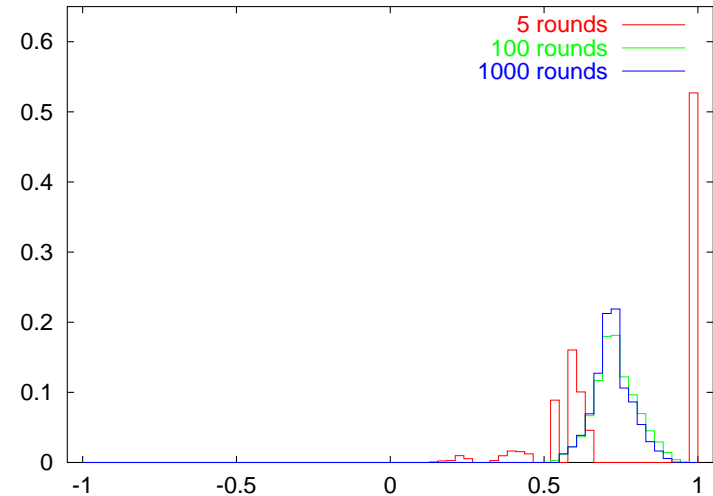
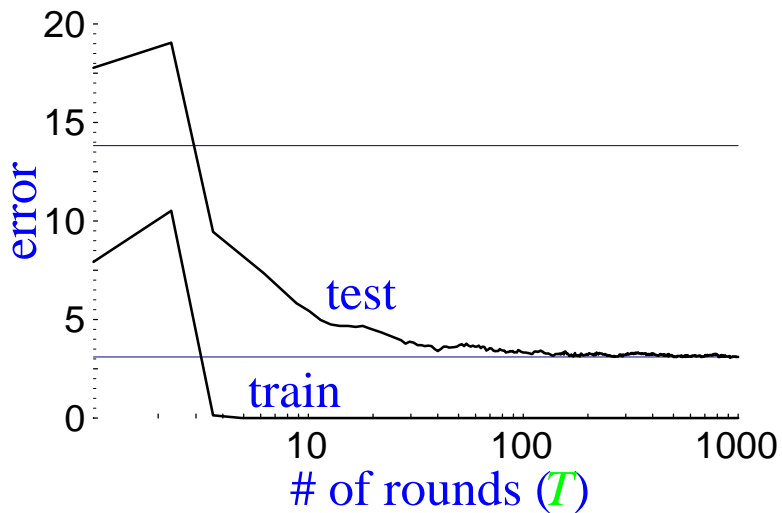
- test error does not increase, even after 1000 rounds
 - (total size > 2,000,000 nodes)
- test error continues to drop even after training error is zero!

	# 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

The Margin Distribution

- margin distribution
= 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 ≤ 0.5	7.7	0.0	0.0
minimum margin	0.14	0.52	0.55