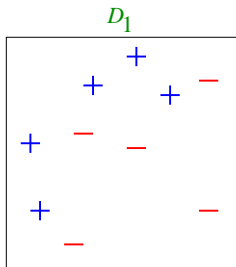
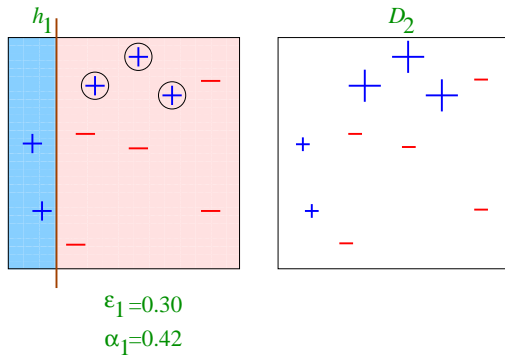


Toy Example

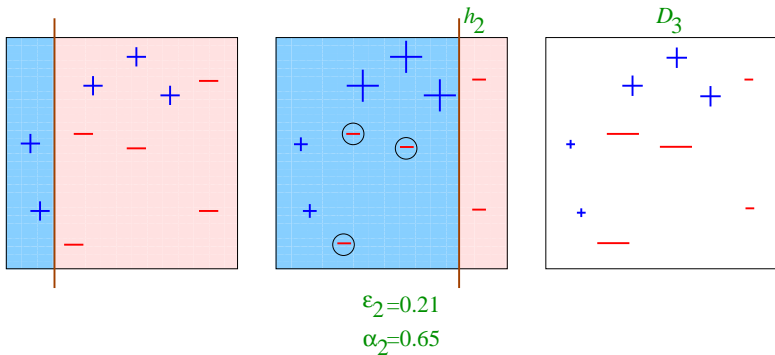


weak hypotheses = vertical or horizontal half-planes

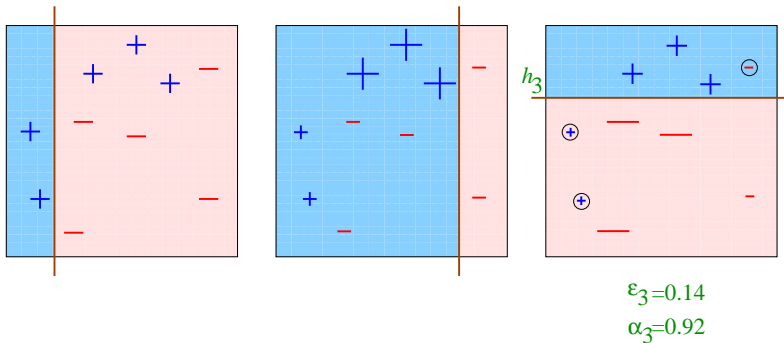
Round 1



Round 2



Round 3



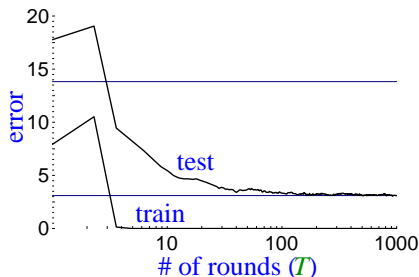
Final Hypothesis

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

$$= \begin{array}{|c|c|c|} \hline \text{blue} & \text{blue} & \text{red} \\ \hline \text{blue} & \text{red} & \text{red} \\ \hline \text{blue} & \text{red} & \text{red} \\ \hline \end{array}$$

The diagram illustrates the final hypothesis H_{final} as a linear combination of three weak hypotheses. Each weak hypothesis is represented by a square with a vertical decision boundary. The first hypothesis has a weight of 0.42 and a vertical boundary at approximately 1/4, with the region to the left being blue and the region to the right being red. The second hypothesis has a weight of 0.65 and a vertical boundary at approximately 3/4, with the region to the left being blue and the region to the right being red. The third hypothesis has a weight of 0.92 and a horizontal boundary at approximately 1/2, with the region above being blue and the region below being red. The final hypothesis is the result of combining these three, shown as a square with two vertical boundaries at approximately 1/4 and 3/4, and one horizontal boundary at approximately 1/2. The regions are colored blue or red, and contain blue '+' signs or red '-' signs to indicate the predicted class for that region.

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!

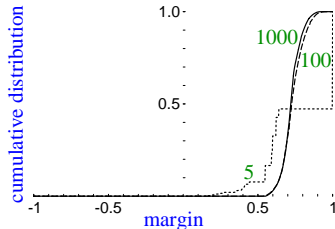
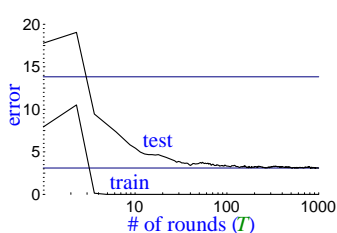
	# rounds		
	5	100	1000
train error	0.0	0.0	0.0
test error	8.4	3.3	3.1

- Occam's razor **wrongly** predicts "simpler" rule is better

The Margin Distribution

- margin distribution

= 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 ≤ 0.5	7.7	0.0	0.0
minimum margin	0.14	0.52	0.55

Application: Detecting Faces

[Viola & Jones]

- **problem**: find **faces** in photograph or movie
- **weak hypotheses**: detect light/dark rectangles in image



- many clever tricks to make extremely fast and accurate