Boosting - Adaboost

Piyush Rai

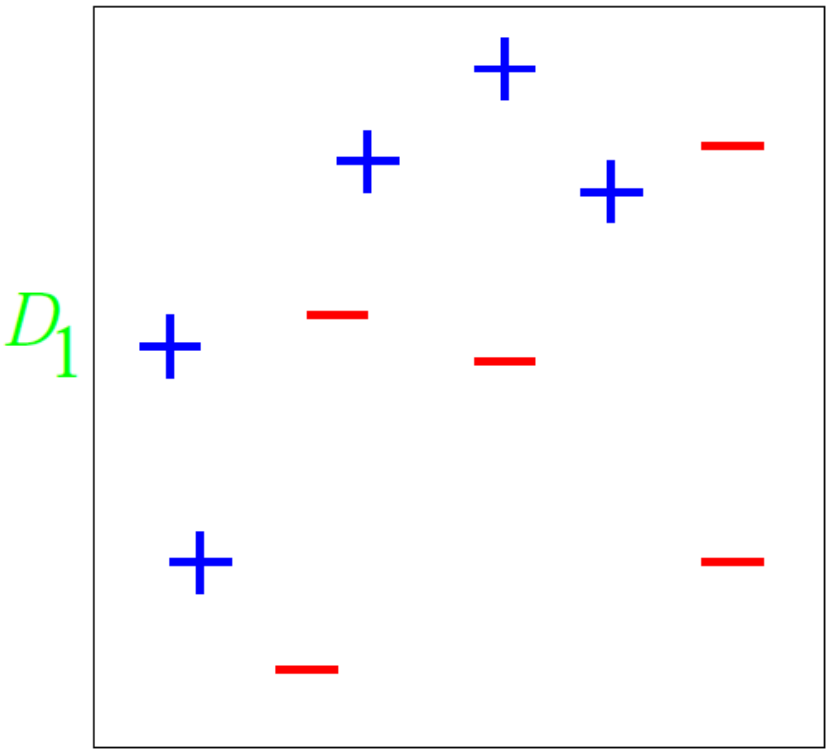
CS5350/6350: Machine Learning

October 27, 2011

Original Dataset

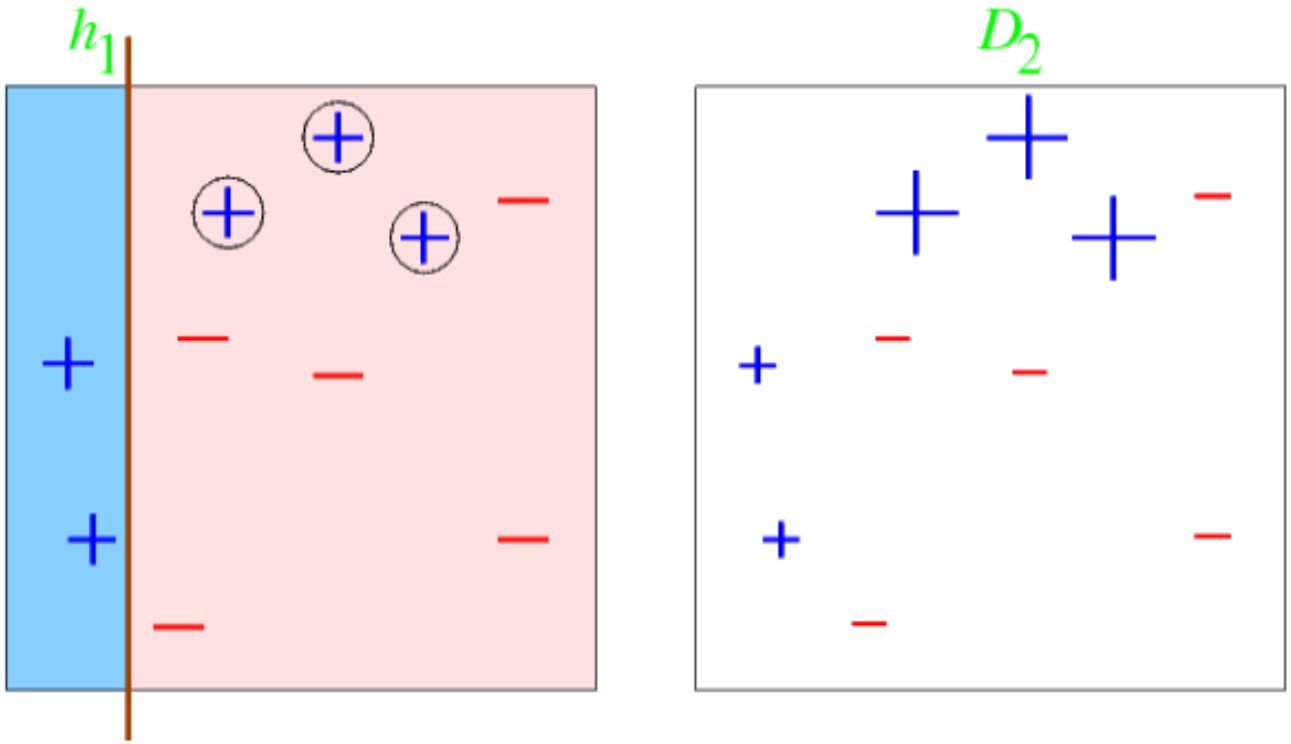
Consider binary classification with 10 training examples

Initial weight distribution D_1 is uniform (each point has equal weight $= 1/10$)



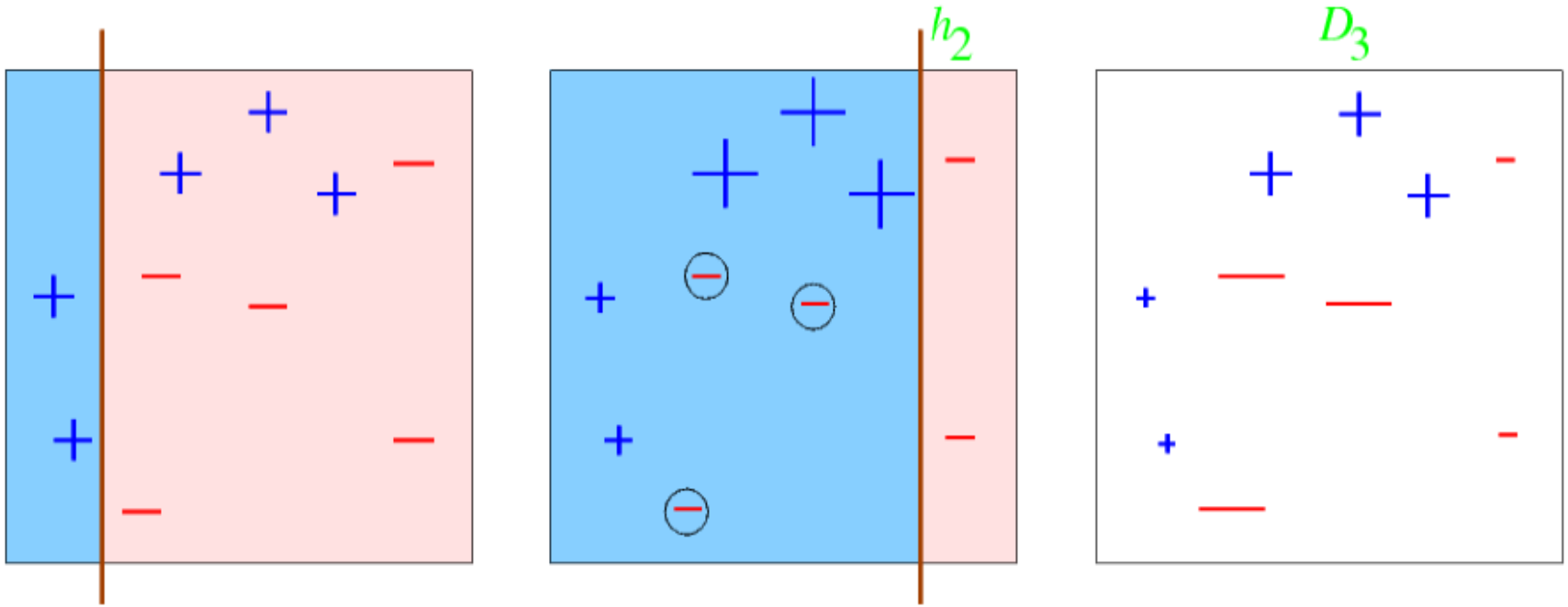
Our weak classifers will be an axis-parallel linear classifiers

After Round 1



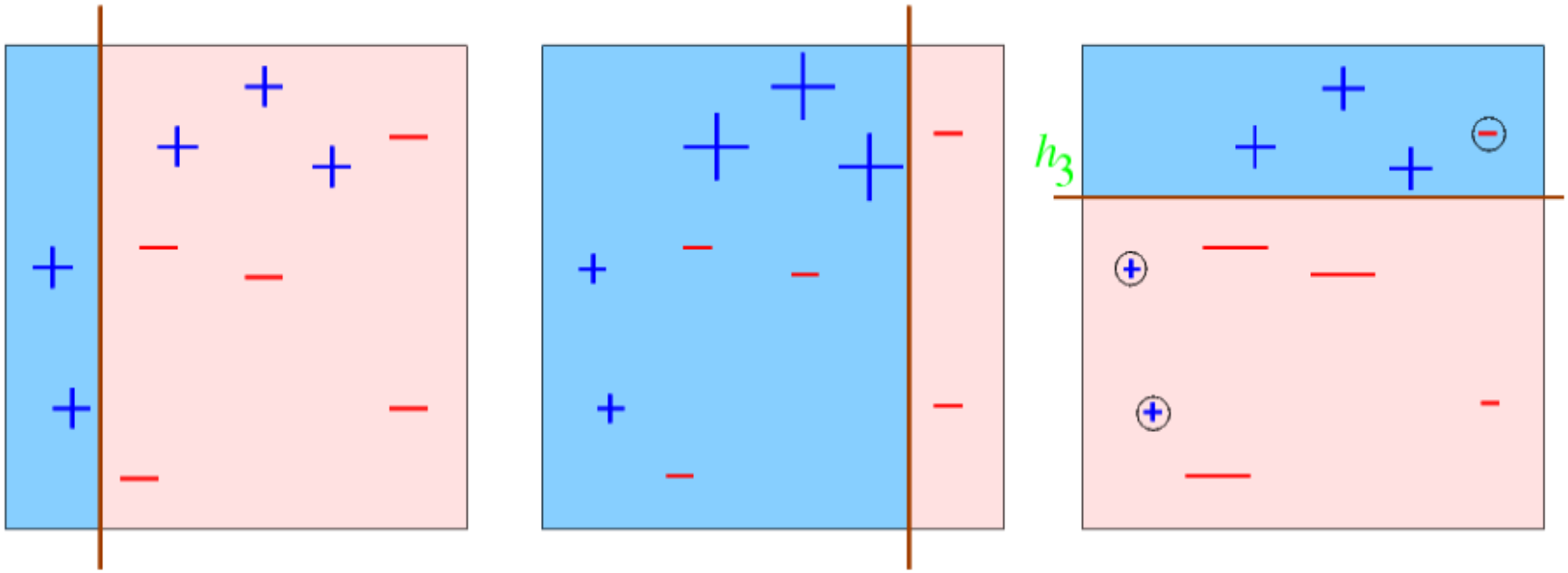
- Error rate of h_1: $\epsilon_1 = 0.3$; weight of h_1: $\alpha_1 = \frac{1}{2}\ln((1-\epsilon_1)/\epsilon_1) = 0.42$
- Each misclassified point upweighted (weight multiplied by e^{α_1})
- Each correctly classified point downweighted (weight multiplied by $e^{-\alpha_1}$)
- Note: Weights are then normalized to 1

After Round 2



- Error rate of h_2: $\epsilon_2 = 0.21$; weight of h_2: $\alpha_2 = \frac{1}{2}\ln((1-\epsilon_2)/\epsilon_2) = 0.65$
- Each misclassified point upweighted (weight multiplied by e^{α_2})
- Each correctly classified point downweighted (weight multiplied by $e^{-\alpha_2}$)
- Note: Weights are then normalized to 1

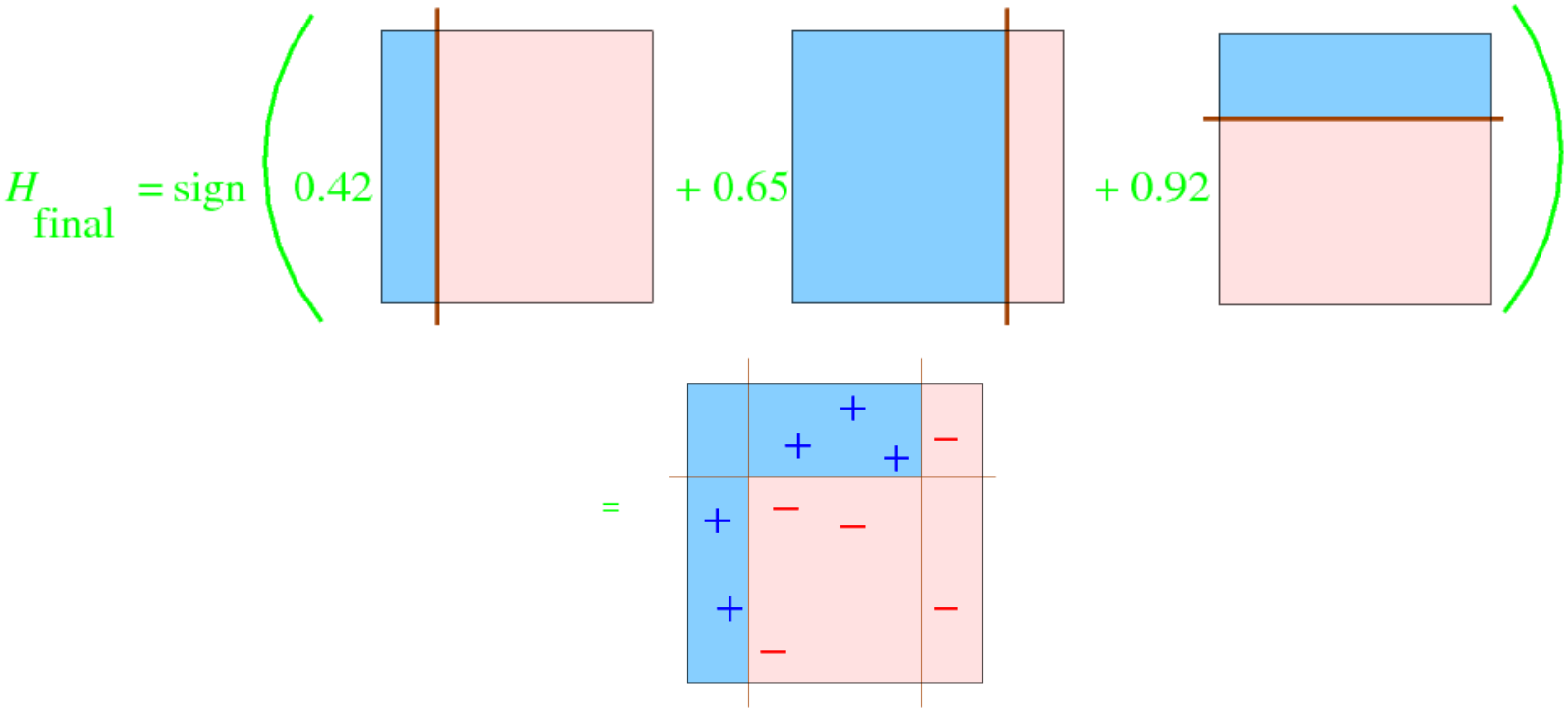
After Round 3



- Error rate of h_3: $\epsilon_3 = 0.14$; weight of h_3: $\alpha_3 = \frac{1}{2}\ln((1-\epsilon_3)/\epsilon_3) = 0.92$
- We decide to stop after round 3
- Our ensemble now consists of 3 classifiers: h_1, h_2, h_3

Final Classifier

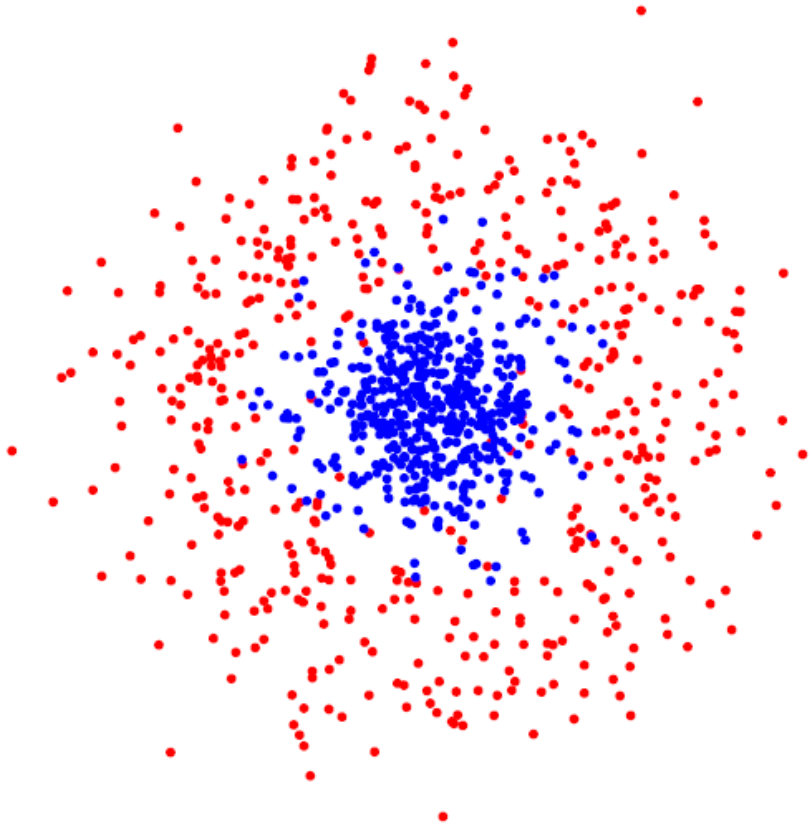
- Final classifier is a weighted linear combination of all the classifiers
- Classifier h_i gets a weight α_i



- Multiple weak, linear classifiers **combined** to give a strong, nonlinear classifier

Another Example

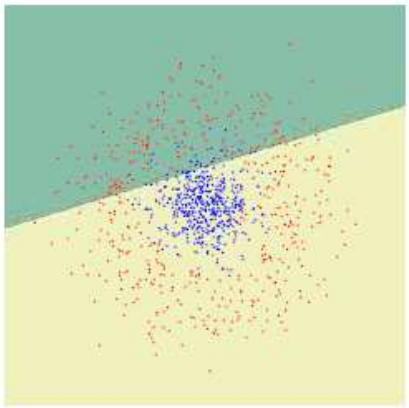
- Given: A nonlinearly separable dataset
- We want to use Perceptron (linear classifier) on this data

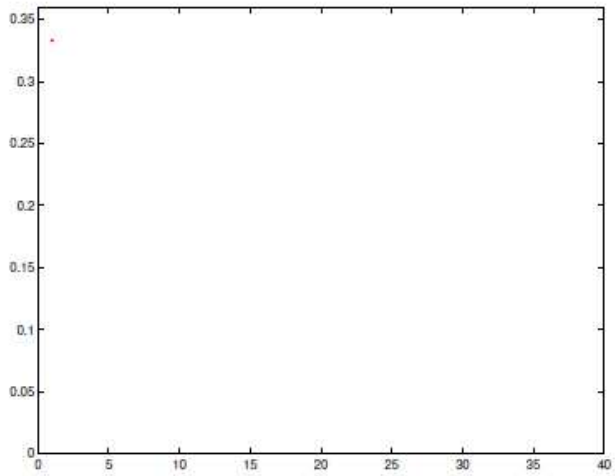


AdaBoost: Round 1

- After round 1, our ensemble has 1 linear classifier (Perceptron)
- Bottom figure: X axis is number of rounds, Y axis is training error

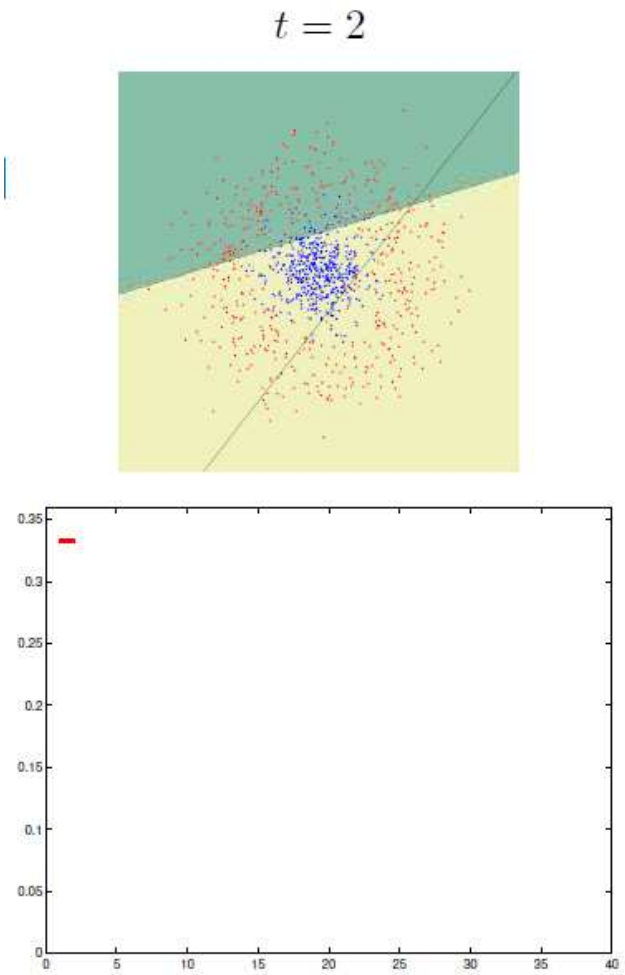
$t = 1$





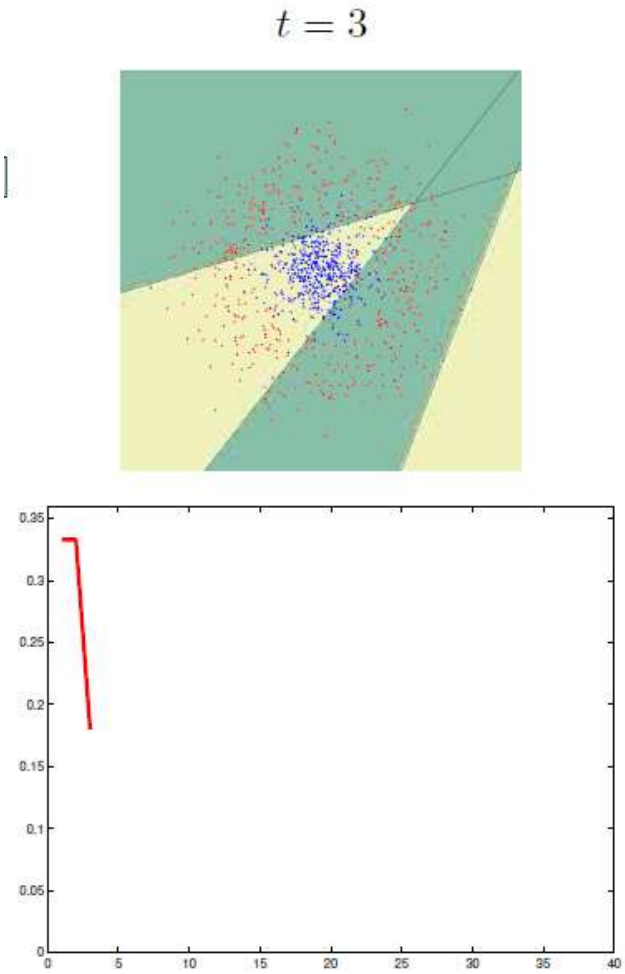
AdaBoost: Round 2

- After round 2, our ensemble has 2 linear classifiers (Perceptrons)
- Bottom figure: X axis is number of rounds, Y axis is training error



AdaBoost: Round 3

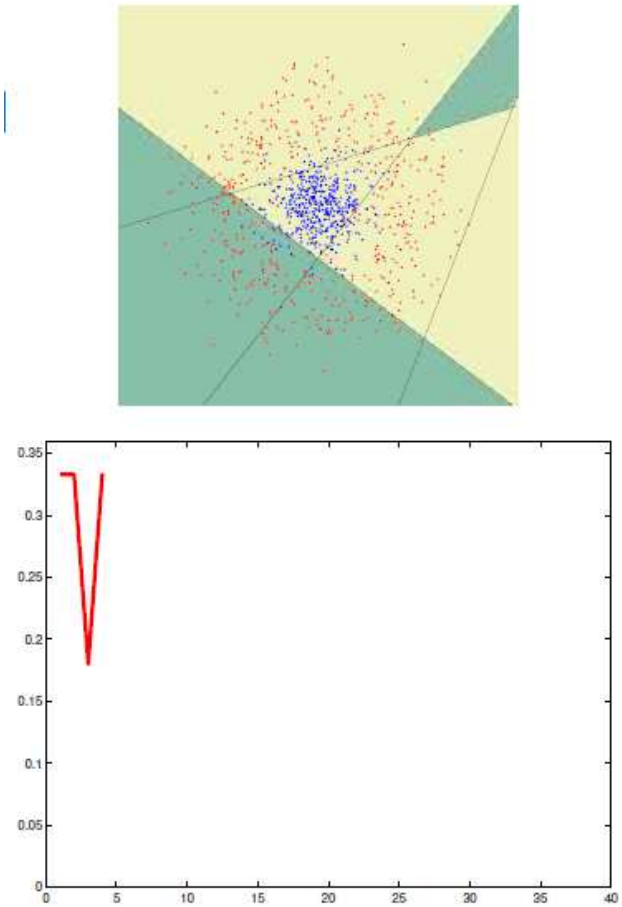
- After round 3, our ensemble has 3 linear classifiers (Perceptrons)
- Bottom figure: X axis is number of rounds, Y axis is training error



AdaBoost: Round 4

- After round 4, our ensemble has 4 linear classifiers (Perceptrons)
- Bottom figure: X axis is number of rounds, Y axis is training error

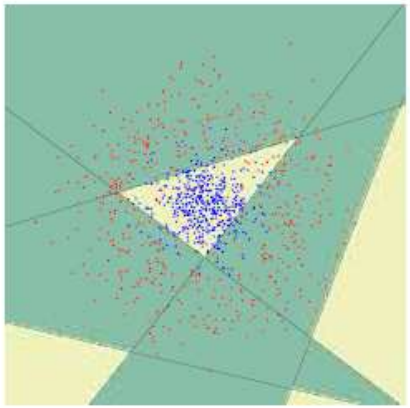
$t = 4$

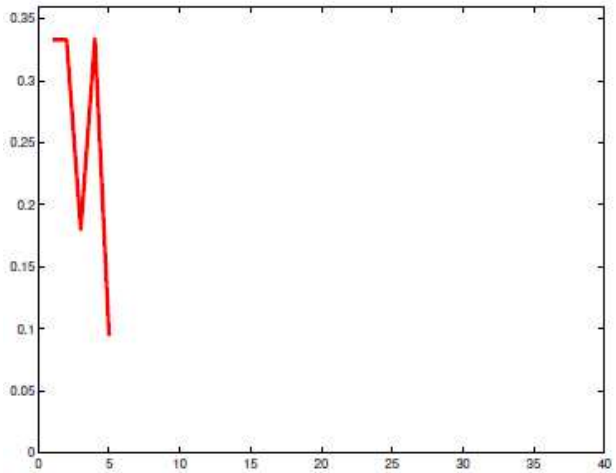


AdaBoost: Round 5

- After round 5, our ensemble has 5 linear classifiers (Perceptrons)
- Bottom figure: X axis is number of rounds, Y axis is training error

$t = 5$

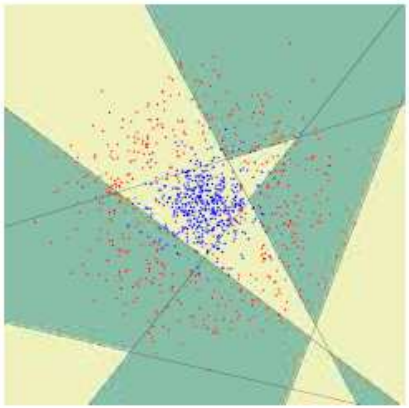


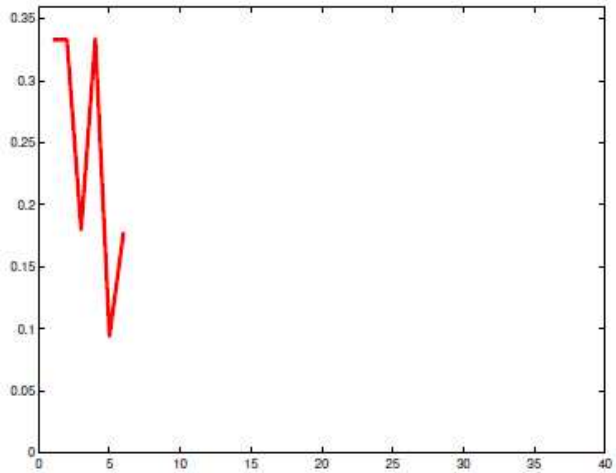


AdaBoost: Round 6

- After round 6, our ensemble has 6 linear classifiers (Perceptrons)
- Bottom figure: X axis is number of rounds, Y axis is training error

$t = 6$

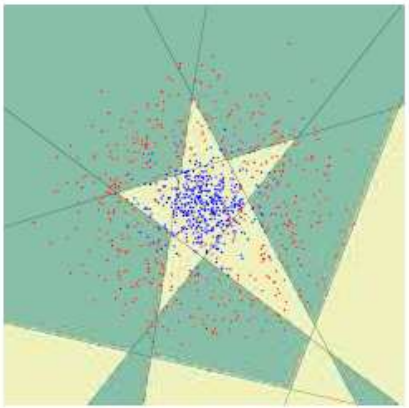


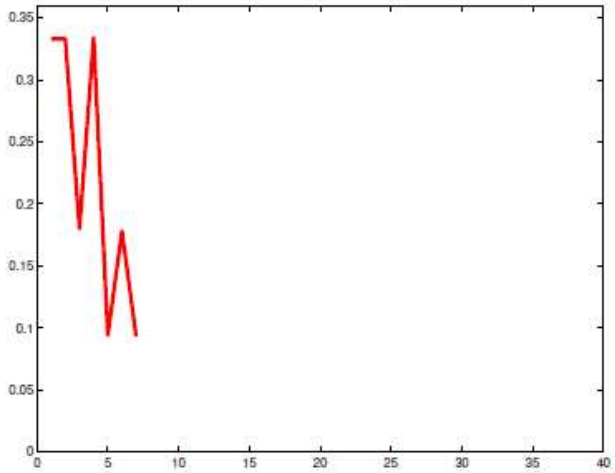


AdaBoost: Round 7

- After round 7, our ensemble has 7 linear classifiers (Perceptrons)
- Bottom figure: X axis is number of rounds, Y axis is training error

$t = 7$





AdaBoost: Round 40

- After round 40, our ensemble has 40 linear classifiers (Perceptrons)
- Bottom figure: X axis is number of rounds, Y axis is training error

$t = 40$