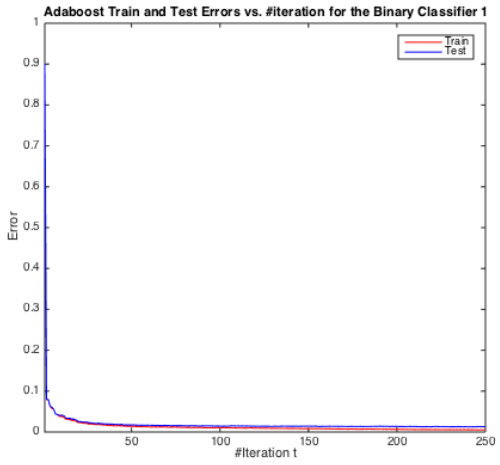
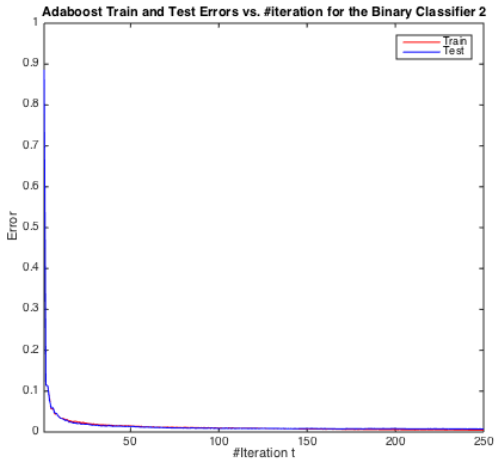
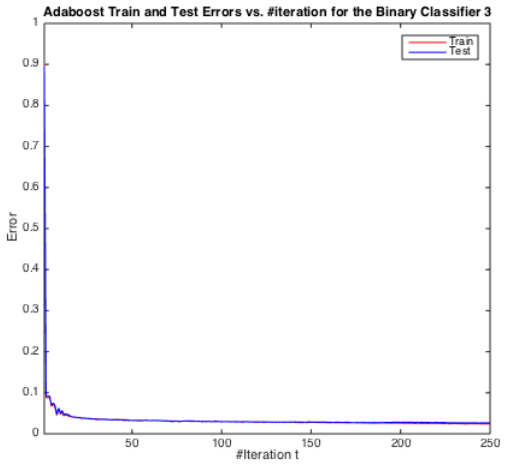
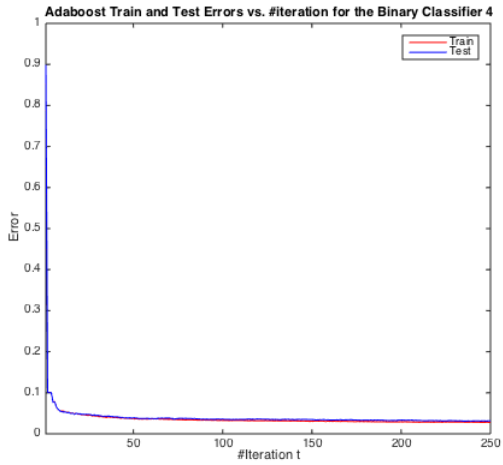
AdaBoost-on-MNIST

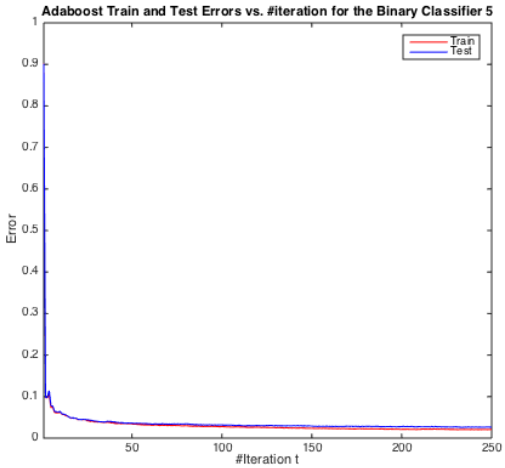
Figure 1 shows the plot of train and test set errors vs. iteration for each binary classifier. The final training and test errors are shown below.

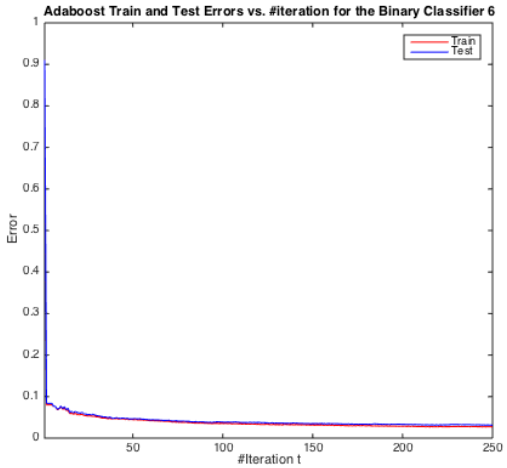












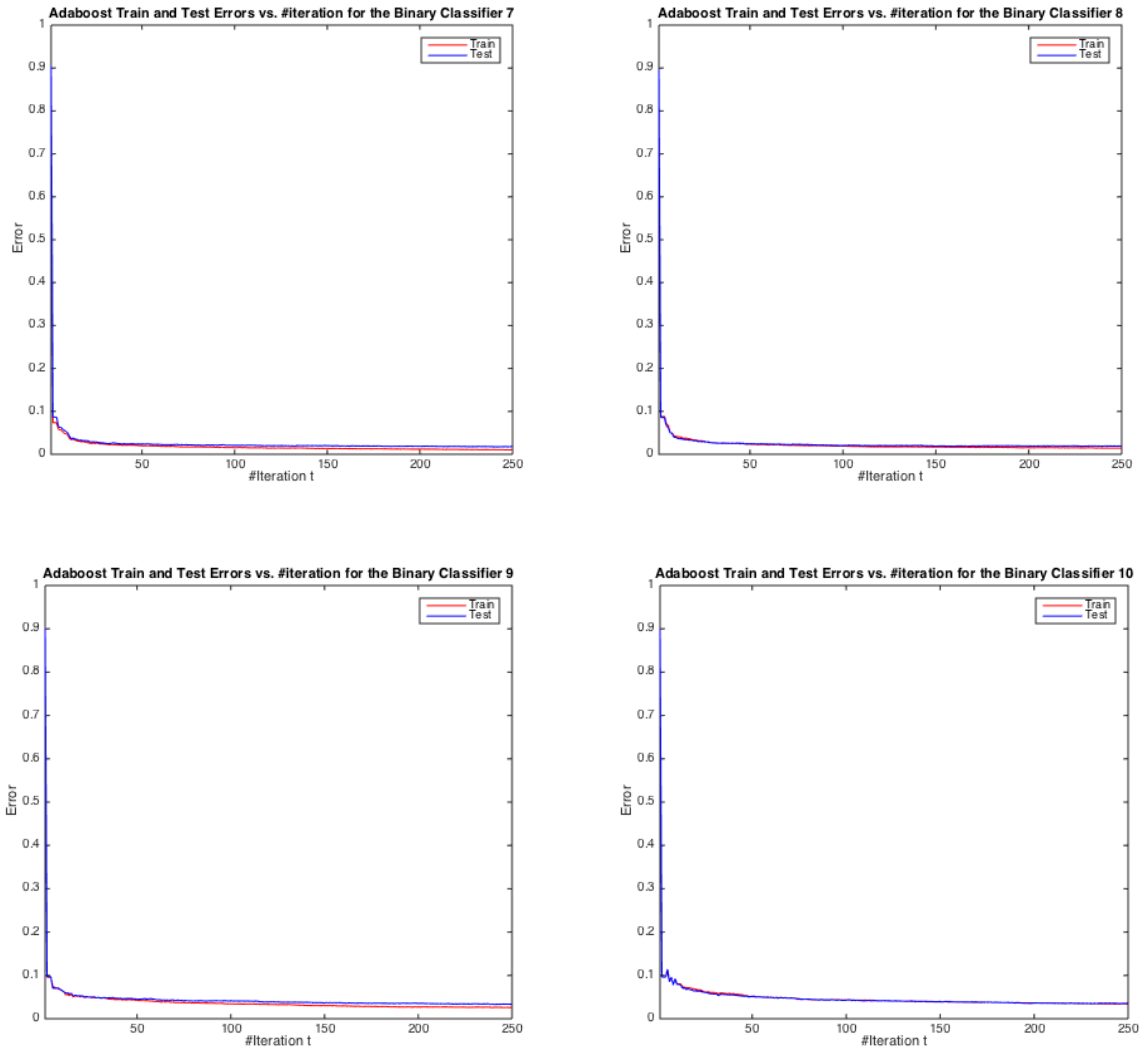
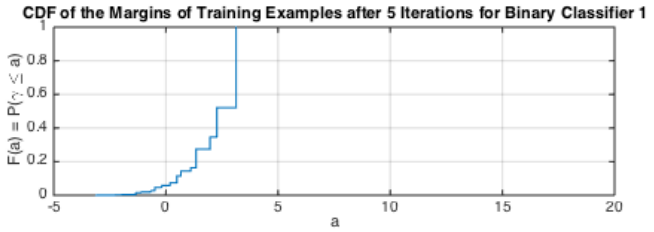


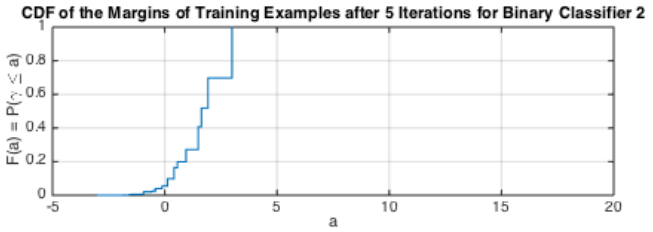
Fig. 1: Training and test errors vs. #iteration for each binary classifier.

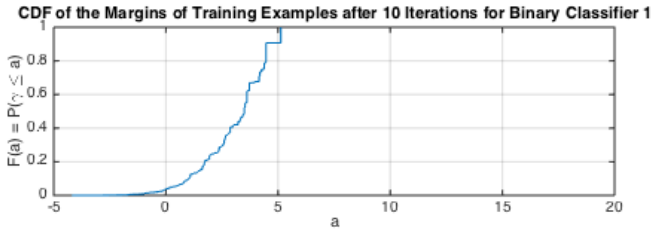
The final training and test error are listed below.

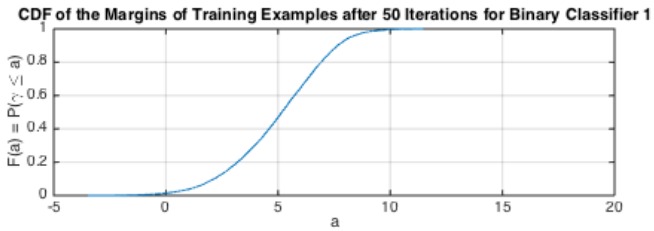
	Final training error	Final test error
Adaboost	0.0648	**0.0920**

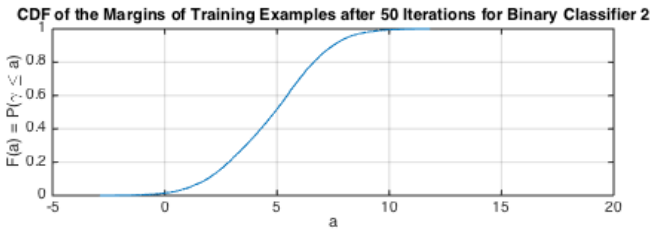
Figure 2 shows the CDF plots of margins for each binary classifier.

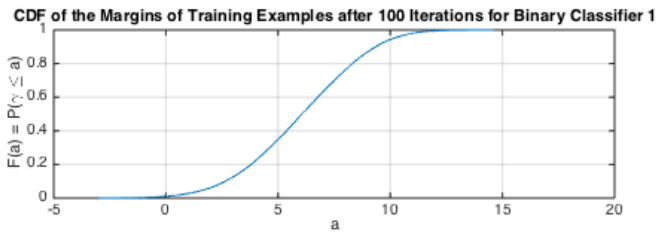


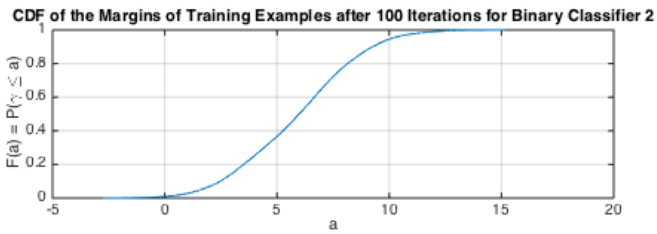


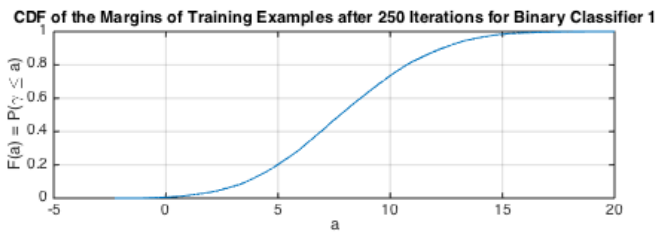


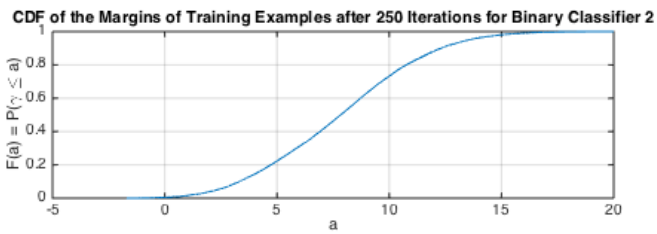


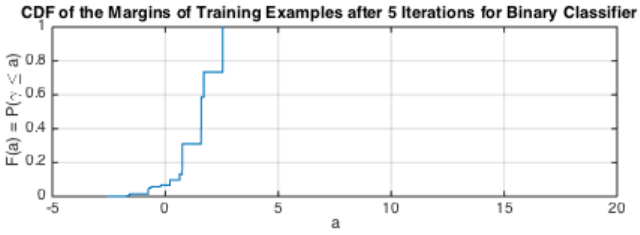


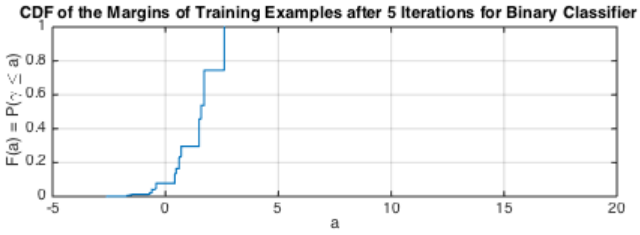


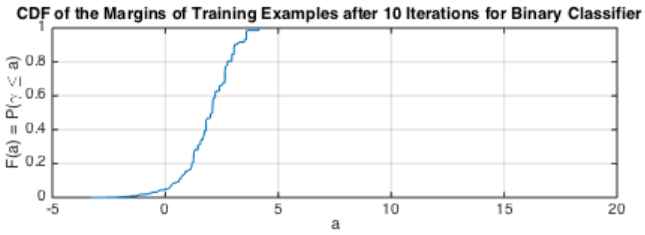


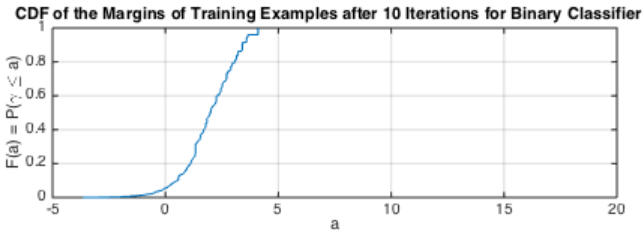


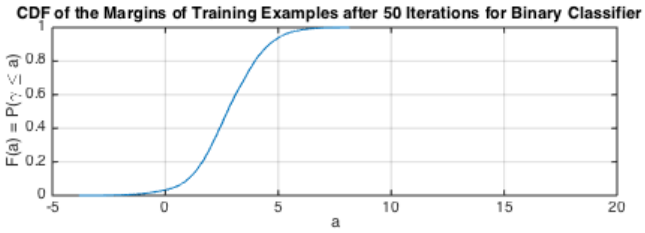


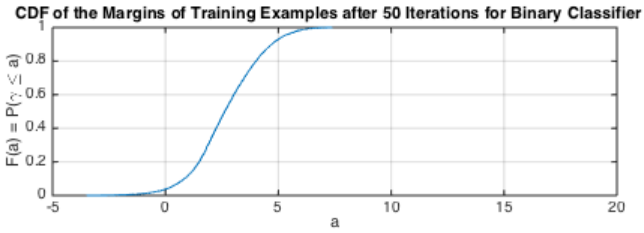

CDF of the Margins of Training Examples after 5 Iterations for Binary Classifier 3
F(a) = P(γ ≤ a)
a

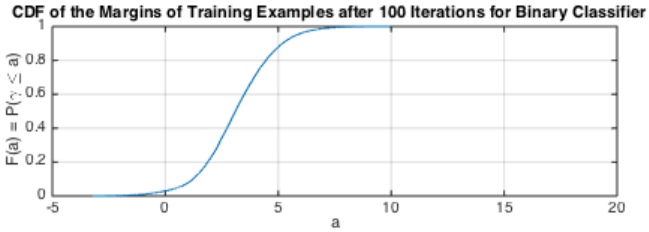

CDF of the Margins of Training Examples after 5 Iterations for Binary Classifier 4
F(a) = P(γ ≤ a)
a

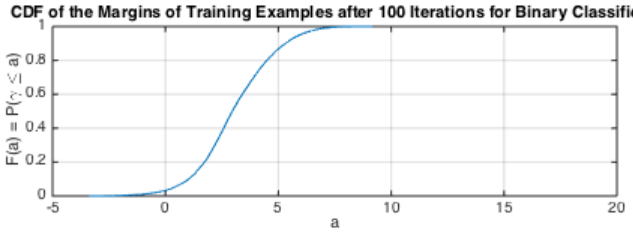

CDF of the Margins of Training Examples after 10 Iterations for Binary Classifier 3
F(a) = P(γ ≤ a)
a

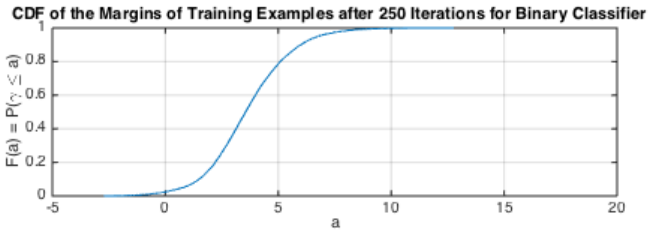

CDF of the Margins of Training Examples after 10 Iterations for Binary Classifier 4
F(a) = P(γ ≤ a)
a

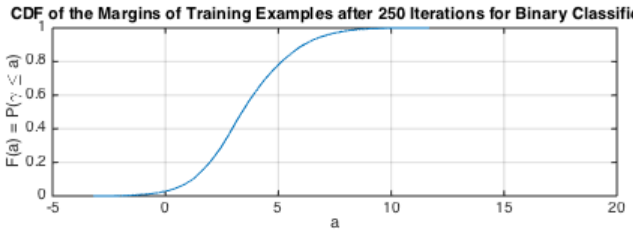

CDF of the Margins of Training Examples after 50 Iterations for Binary Classifier 3
F(a) = P(γ ≤ a)
a

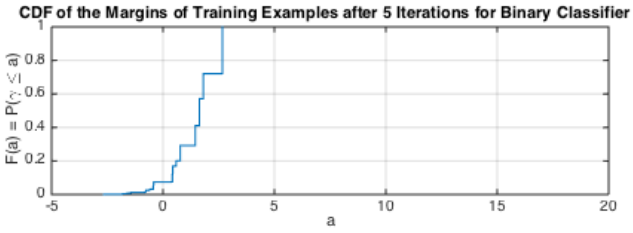

CDF of the Margins of Training Examples after 50 Iterations for Binary Classifier 4
F(a) = P(γ ≤ a)
a

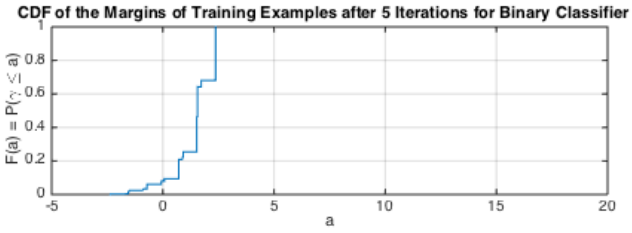

CDF of the Margins of Training Examples after 100 Iterations for Binary Classifier 3
F(a) = P(γ ≤ a)
a

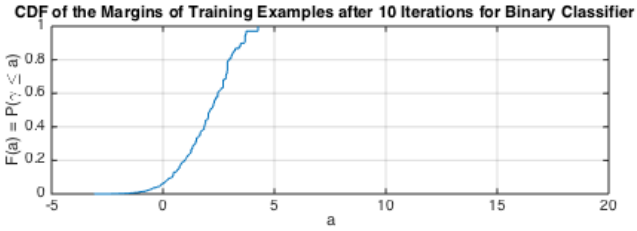

CDF of the Margins of Training Examples after 100 Iterations for Binary Classifier 4
F(a) = P(γ ≤ a)
a

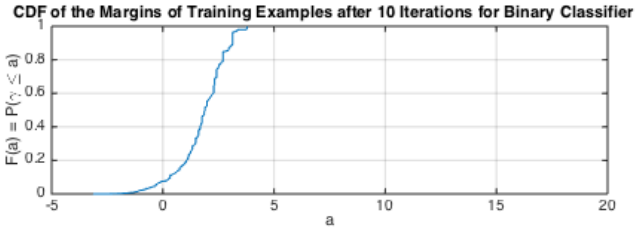

CDF of the Margins of Training Examples after 250 Iterations for Binary Classifier 3
F(a) = P(γ ≤ a)
a

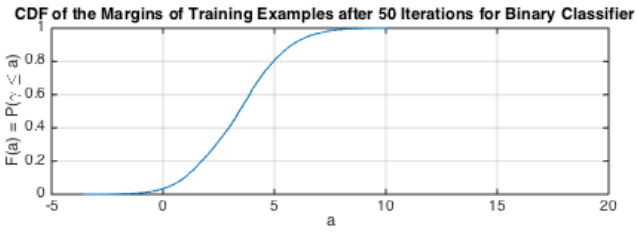

CDF of the Margins of Training Examples after 250 Iterations for Binary Classifier 4
F(a) = P(γ ≤ a)
a

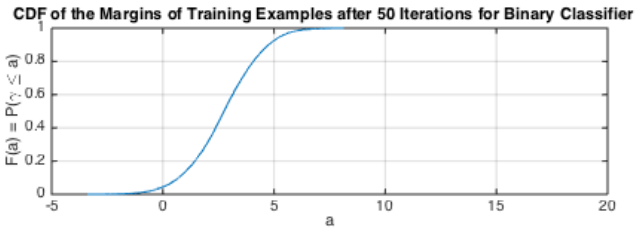

CDF of the Margins of Training Examples after 5 Iterations for Binary Classifier 5
F(a) = P(γ ≤ a)
a

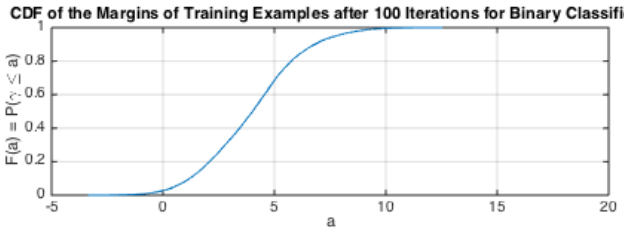

CDF of the Margins of Training Examples after 5 Iterations for Binary Classifier 6
F(a) = P(γ ≤ a)
a

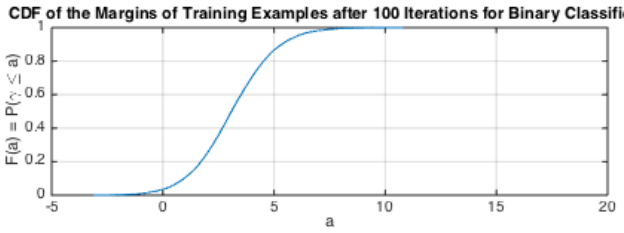

CDF of the Margins of Training Examples after 10 Iterations for Binary Classifier 5
F(a) = P(γ ≤ a)
a

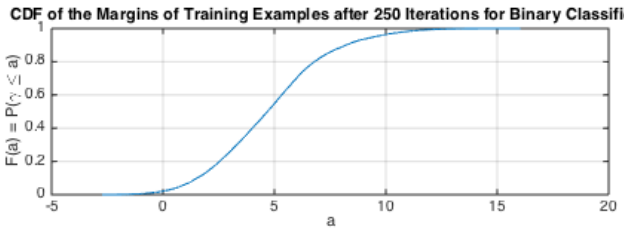

CDF of the Margins of Training Examples after 10 Iterations for Binary Classifier 6
F(a) = P(γ ≤ a)
a

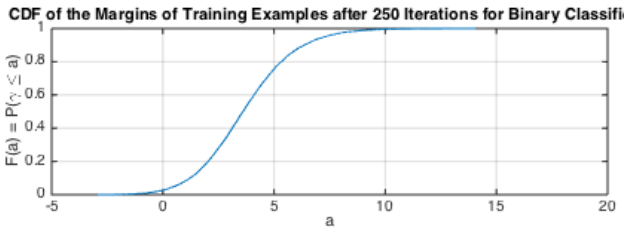

CDF of the Margins of Training Examples after 50 Iterations for Binary Classifier 5
F(a) = P(γ ≤ a)
a

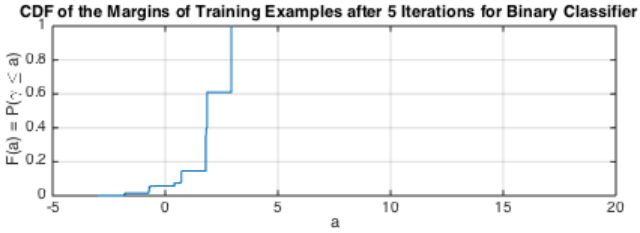

CDF of the Margins of Training Examples after 50 Iterations for Binary Classifier 6
F(a) = P(γ ≤ a)
a

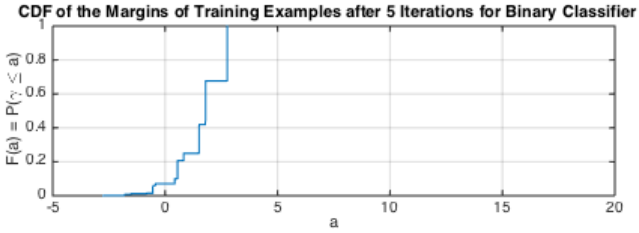

CDF of the Margins of Training Examples after 100 Iterations for Binary Classifier 5
F(a) = P(γ ≤ a)
a

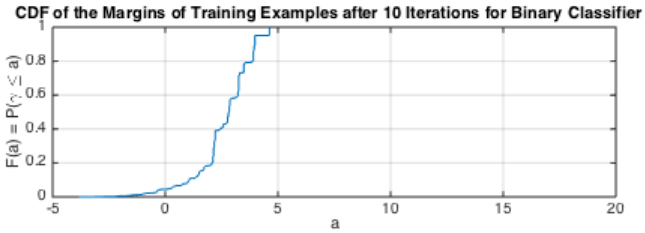

CDF of the Margins of Training Examples after 100 Iterations for Binary Classifier 6
F(a) = P(γ ≤ a)
a

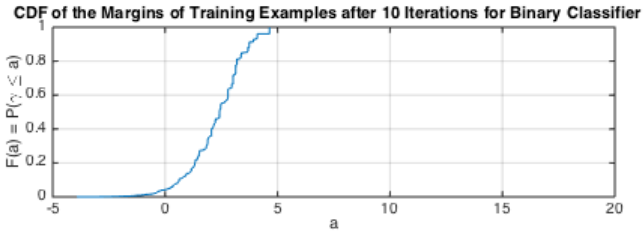

CDF of the Margins of Training Examples after 250 Iterations for Binary Classifier 5
F(a) = P(γ ≤ a)
a

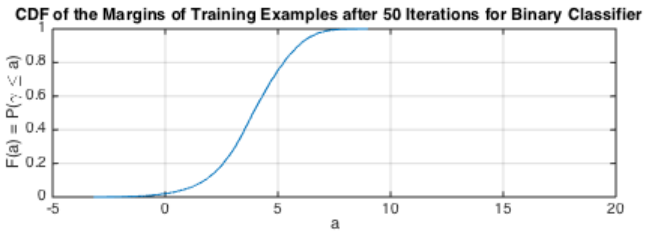

CDF of the Margins of Training Examples after 250 Iterations for Binary Classifier 6
F(a) = P(γ ≤ a)
a

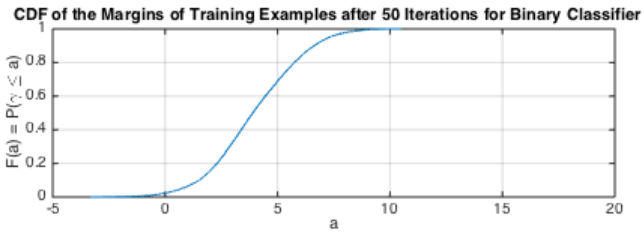

CDF of the Margins of Training Examples after 5 Iterations for Binary Classifier 7
F(a) = P(γ ≤ a)
a

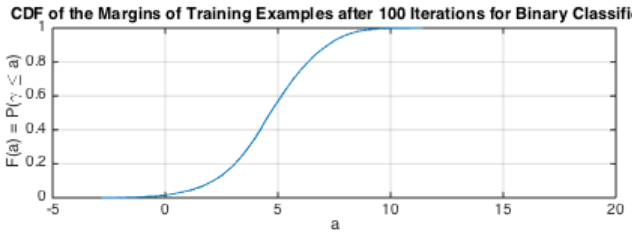

CDF of the Margins of Training Examples after 5 Iterations for Binary Classifier 8
F(a) = P(γ ≤ a)
a

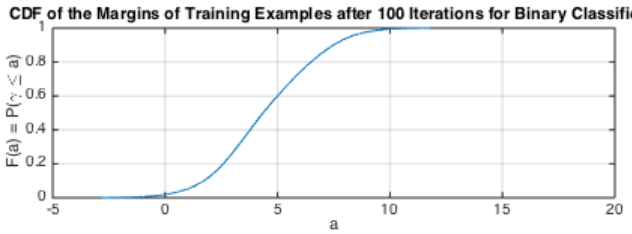

CDF of the Margins of Training Examples after 10 Iterations for Binary Classifier 7
F(a) = P(γ ≤ a)
a

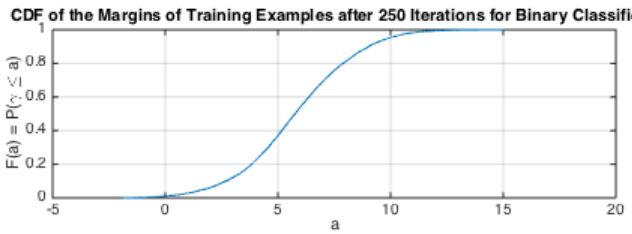

CDF of the Margins of Training Examples after 10 Iterations for Binary Classifier 8
F(a) = P(γ ≤ a)
a

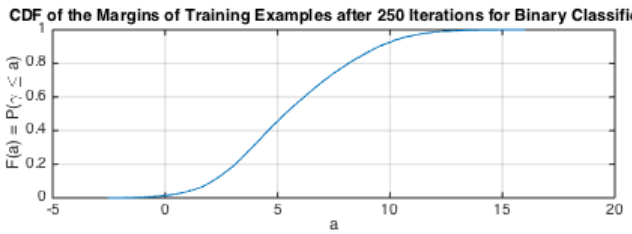

CDF of the Margins of Training Examples after 50 Iterations for Binary Classifier 7
F(a) = P(γ ≤ a)
a


CDF of the Margins of Training Examples after 50 Iterations for Binary Classifier 8
F(a) = P(γ ≤ a)
a


CDF of the Margins of Training Examples after 100 Iterations for Binary Classifier 7
F(a) = P(γ ≤ a)
a


CDF of the Margins of Training Examples after 100 Iterations for Binary Classifier 8
F(a) = P(γ ≤ a)
a


CDF of the Margins of Training Examples after 250 Iterations for Binary Classifier 7
F(a) = P(γ ≤ a)
a


CDF of the Margins of Training Examples after 250 Iterations for Binary Classifier 8
F(a) = P(γ ≤ a)
a

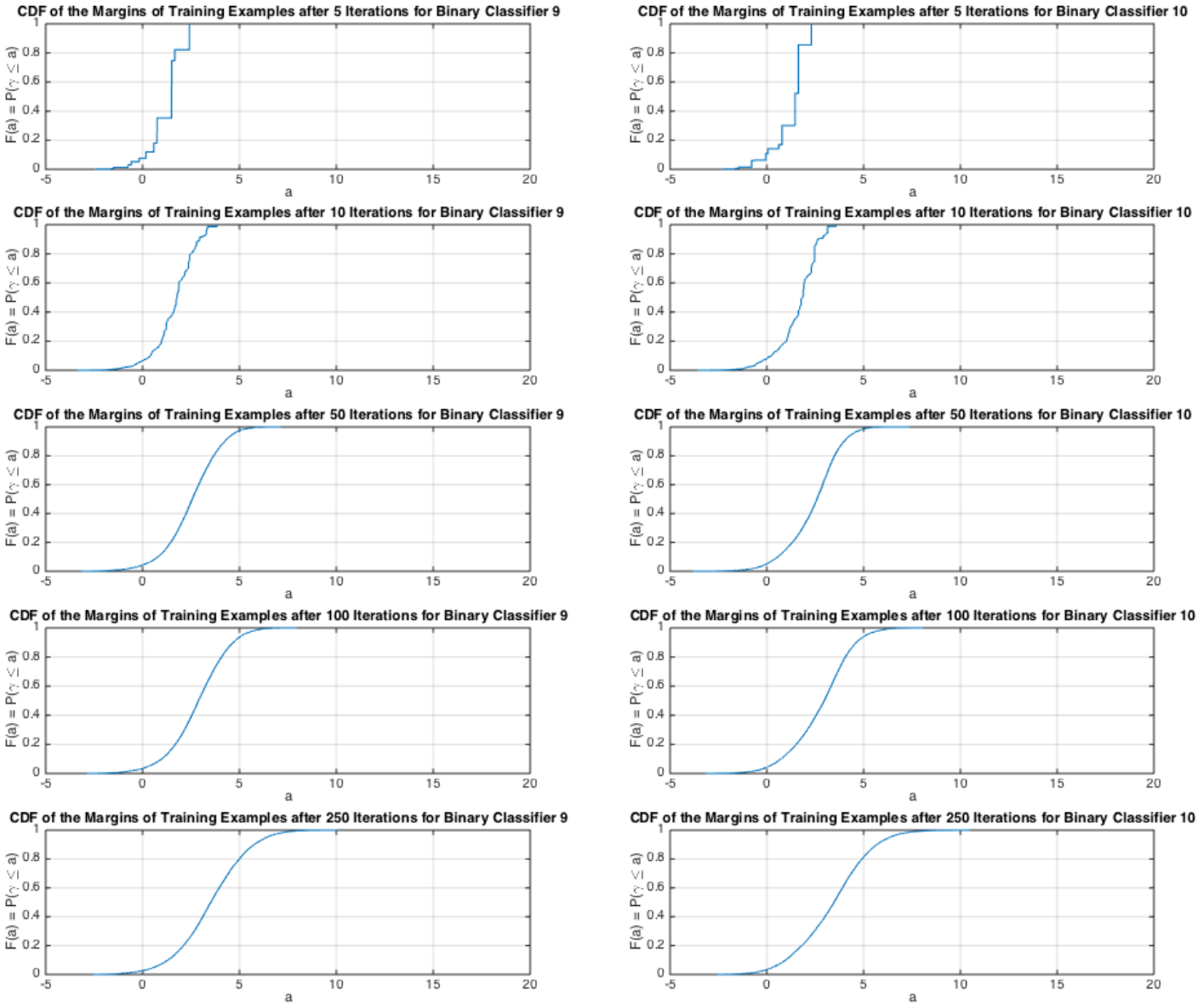
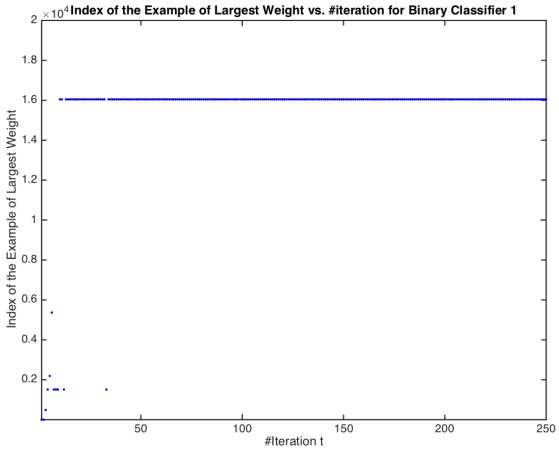
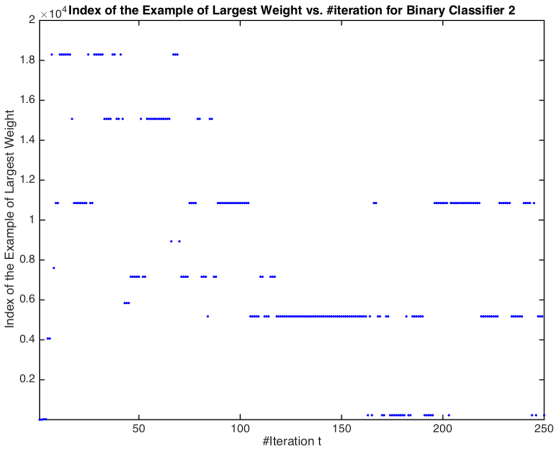


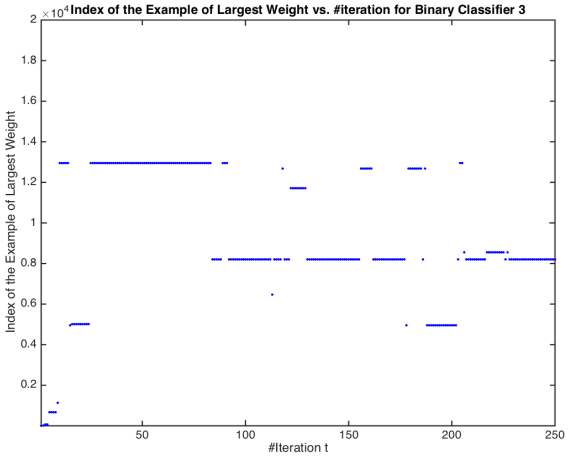
Fig. 2: CDF of the margins vs. #iteration for each binary classifier.

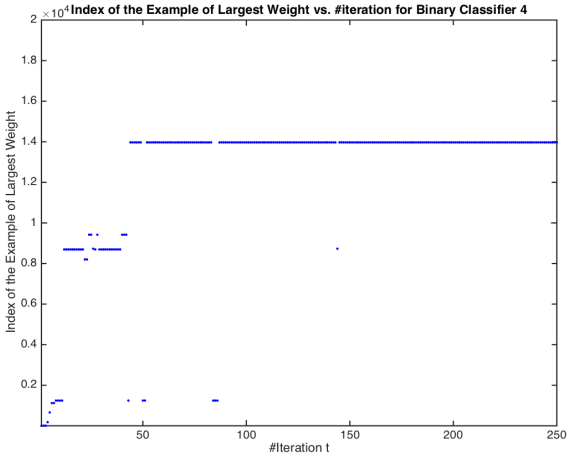
The results show that during the increasing of boosting iterations, large margins get higher probability, which means that boosting can encourage large margins.

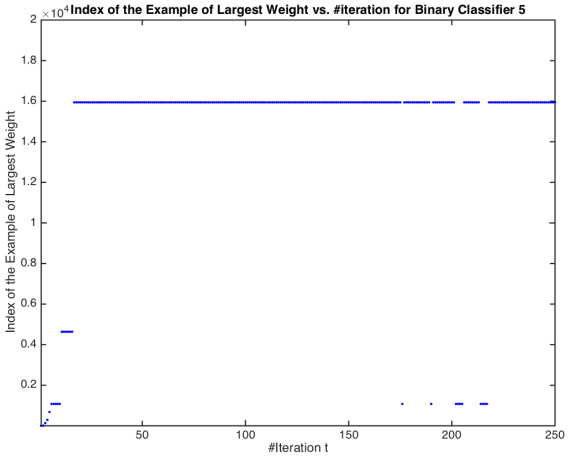
Figure 3 and Figure 4 plot the index of the examples of largest weight for each each boosting iteration and the top 3 “heaviest” examples.

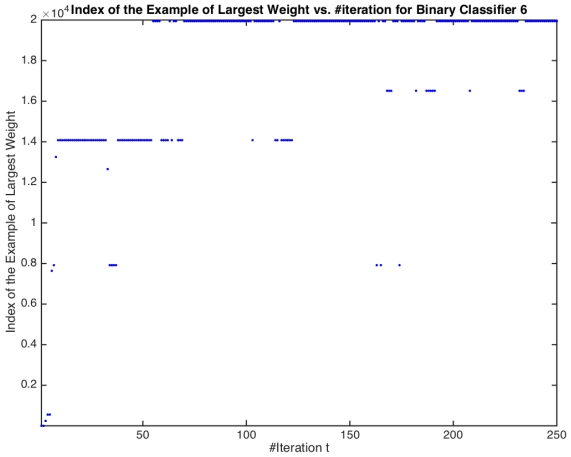












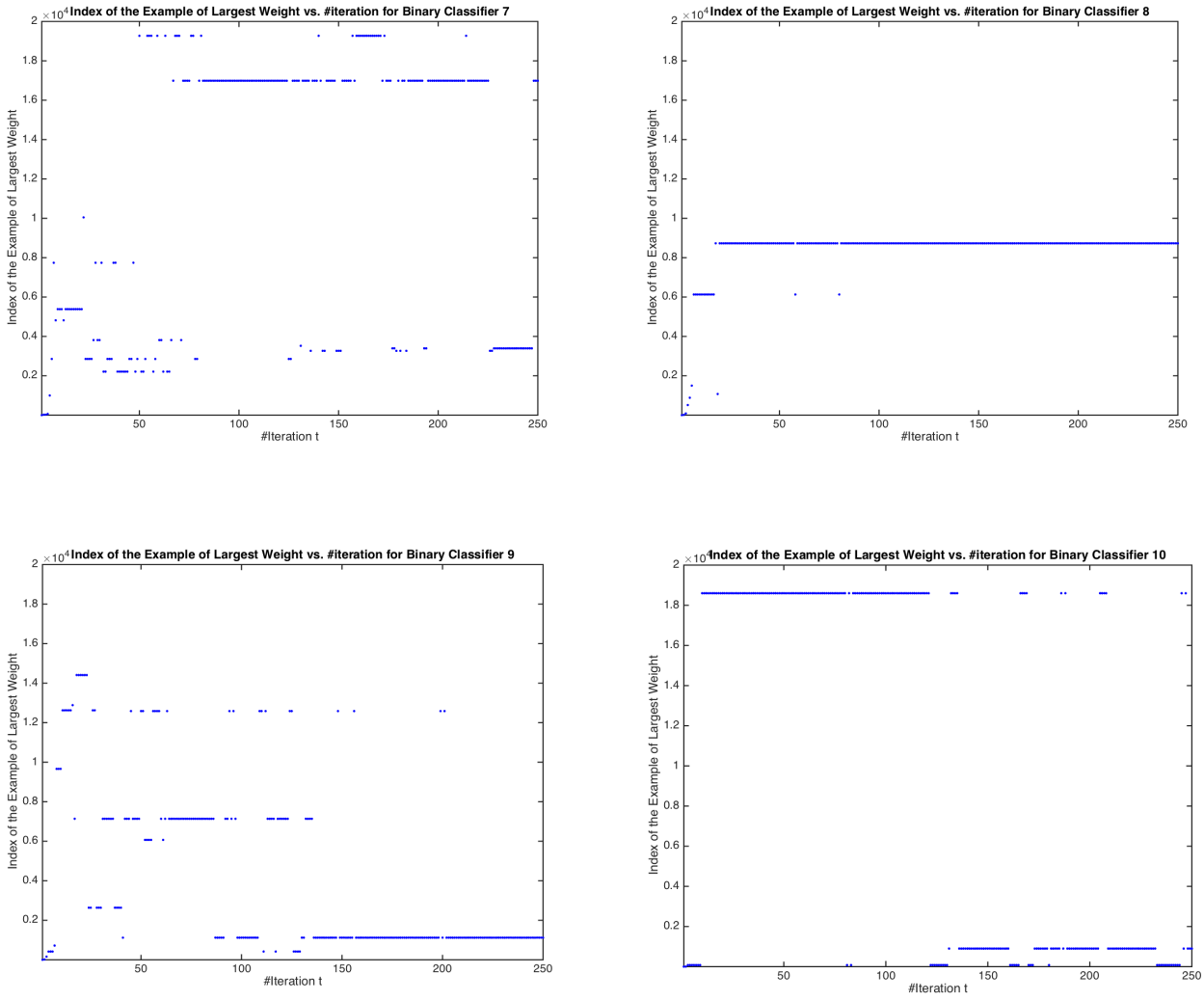
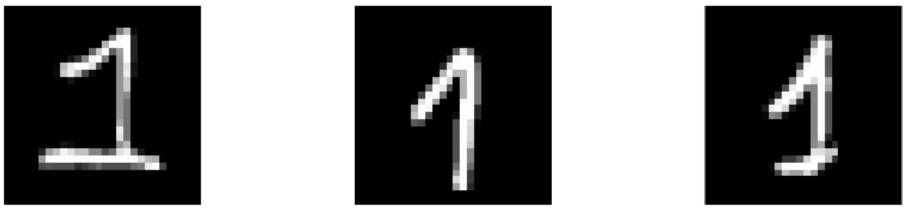
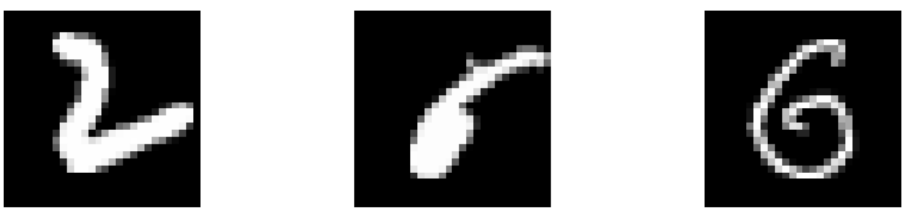


Fig. 3: The index of the example of largest weight vs. #iteration for each binary classifier.





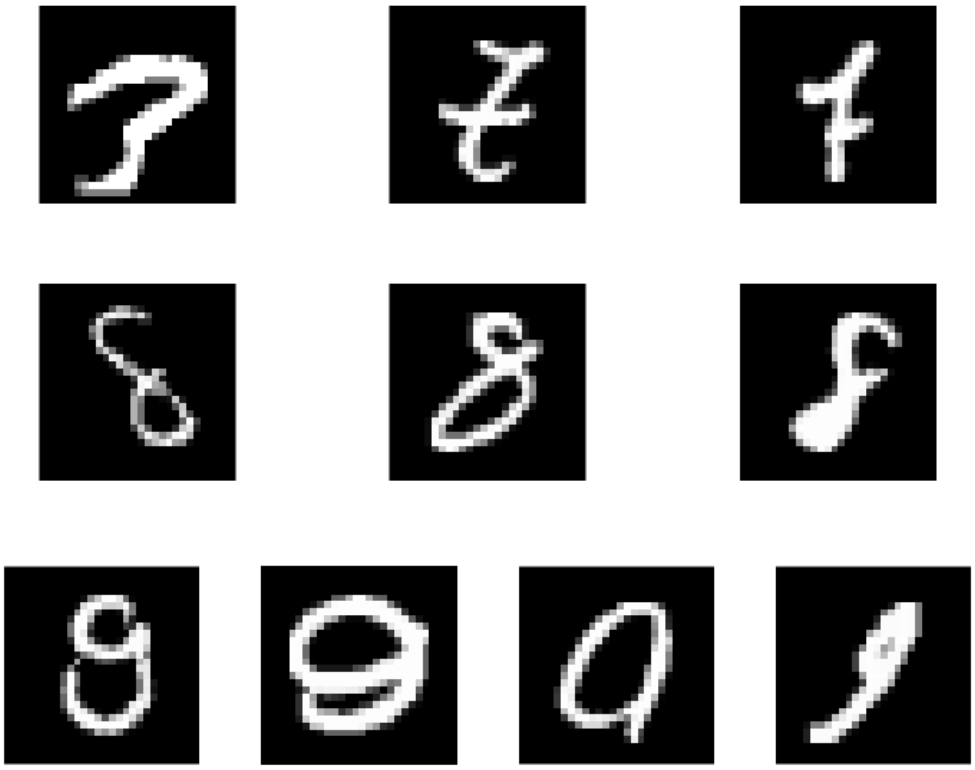
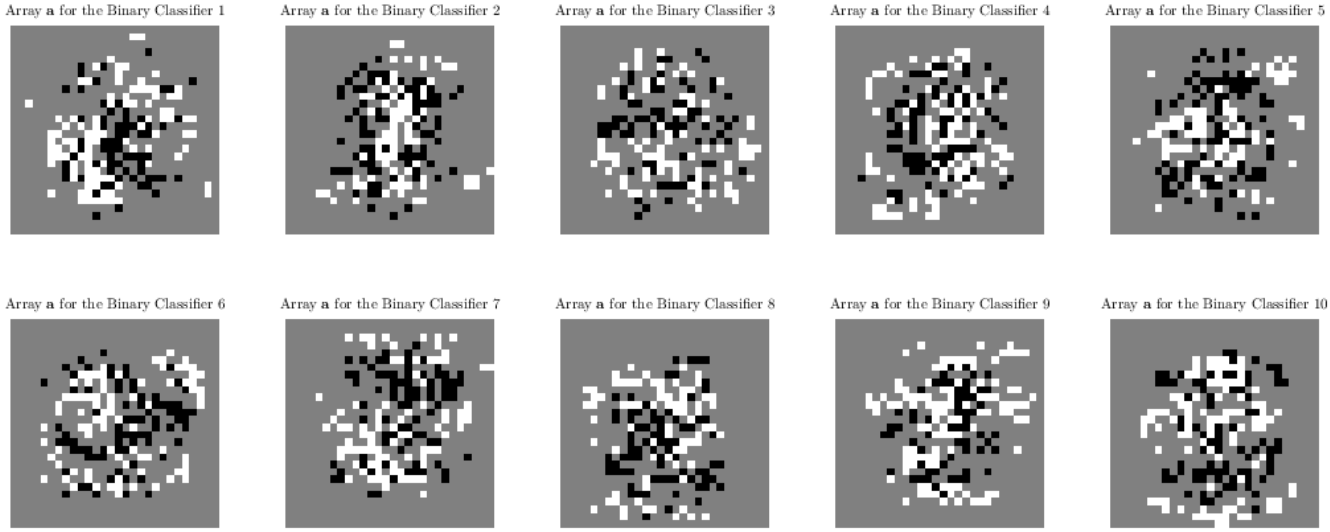


Fig. 4: Top 3 “heaviest” examples for each binary classifier.

From the plots above we can conclude that the weighting mechanism makes boosting focus on hard examples. They are either false positives or false negatives.

Figure 5 visualizes the weaker learners for each binary classifier.



The visualization above indicates that what weaker classifiers are doing is actually a feature selection process. The selected features tend to fall in the same locations corresponding to each hand-written digits.