

Fig.1 shows the plots of the train and test error rates vs. iteration for binary classifiers. The error rates start from about 10% and decrease as iterations proceed.

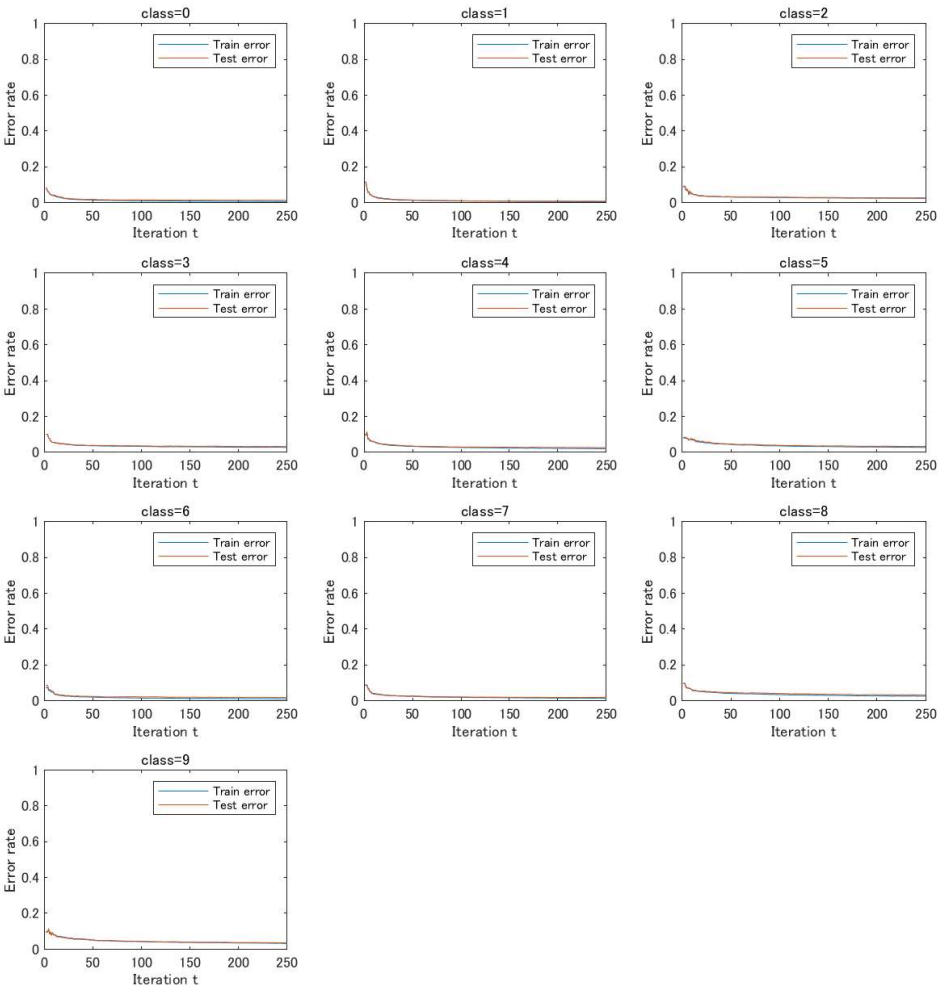


Fig.1 the train and test error rates vs. iteration for binary classifiers.

The Fig.2 shows the test error rate of the multi-class classifier. The final error rate is 0.0936.

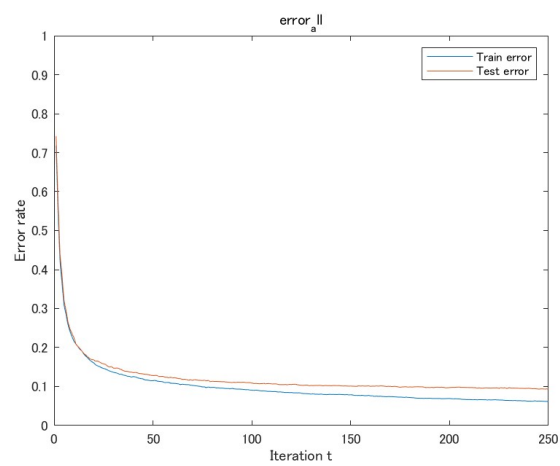


Fig.1 the train and test error rates vs. iteration for the multi-class classifier.

The Fig.3 shows the cumulative distribution function (CDF) for of the margins of all training samples after $\{5, 10, 50, 100, 250\}$ iterations.

As iterations proceed, the CDF lines shift from left to right, and the range of the margins becomes greater. This means that the classifiers enforce the margins. Since the exponential loss function is used for this algorithm, the positive margins are also enforced.

The intersections of the DCF lines and the vertical axis are the error rates of the binary classifiers at the iteration. As iterations proceed, the intersection points go down, which means that the binary classifiers improve the accuracy for training data.

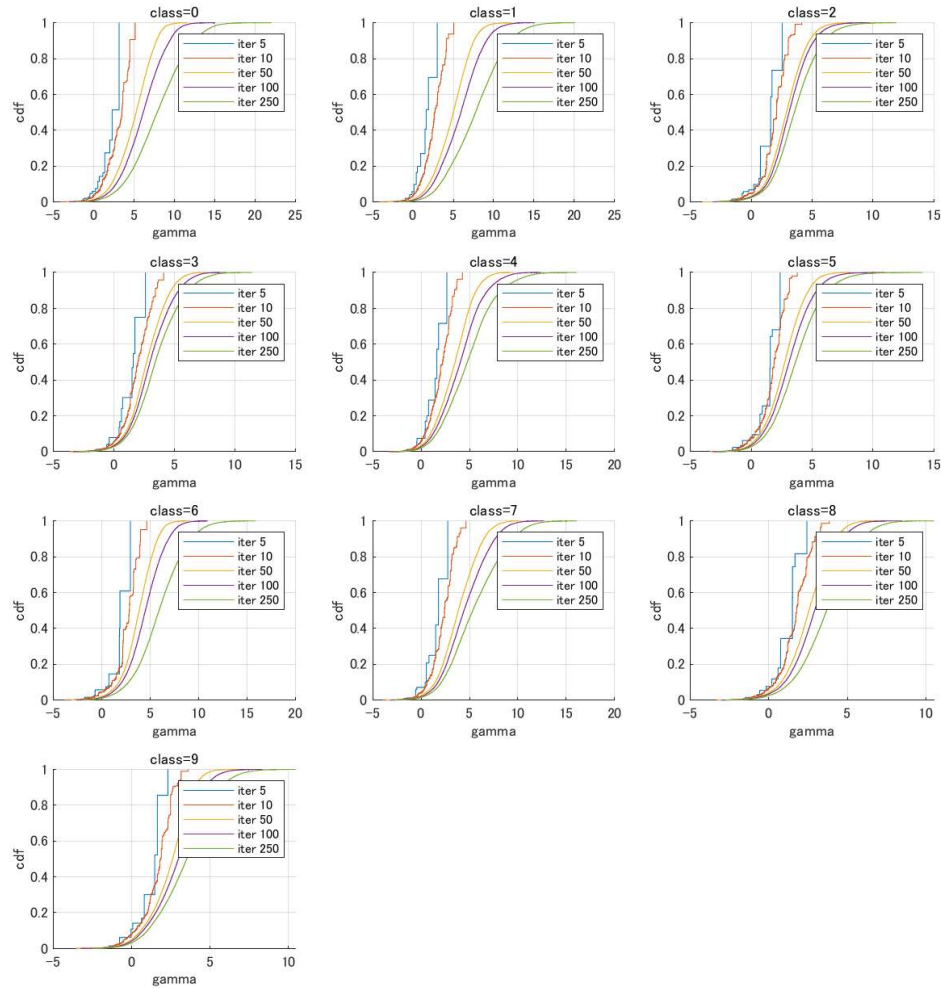


Fig.3 the cumulative distribution function for of the margins of all training samples after $\{5, 10, 50, 100, 250\}$ iterations

The Fig.4 shows that the index of the example of largest weight for each boosting iteration. There are some difficult images to be classified.

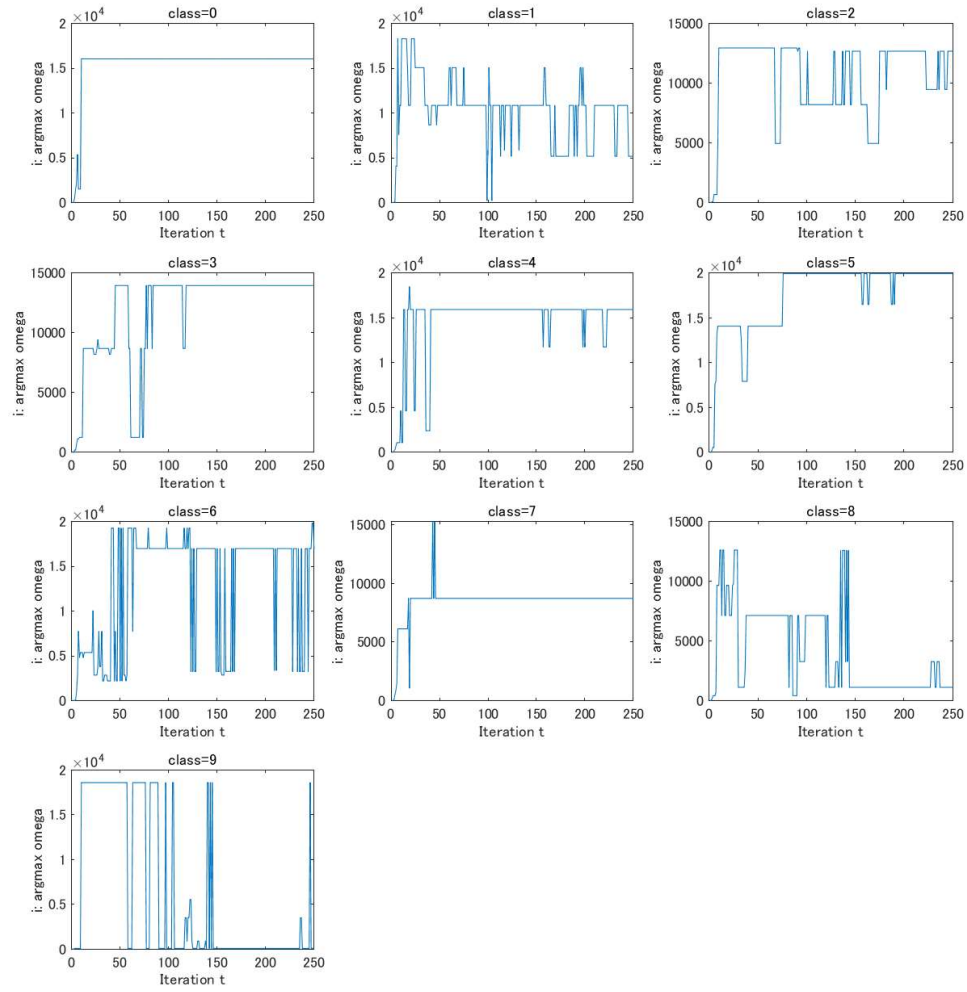


Fig.4 The index of the example of largest weight for each boosting iteration.

The Fig.5 shows the 3 examples that were most frequently selected, across boosting iterations, as the example of largest weight. According to the Fig.4, the left image of class 0 and class 7 are especially hard examples.

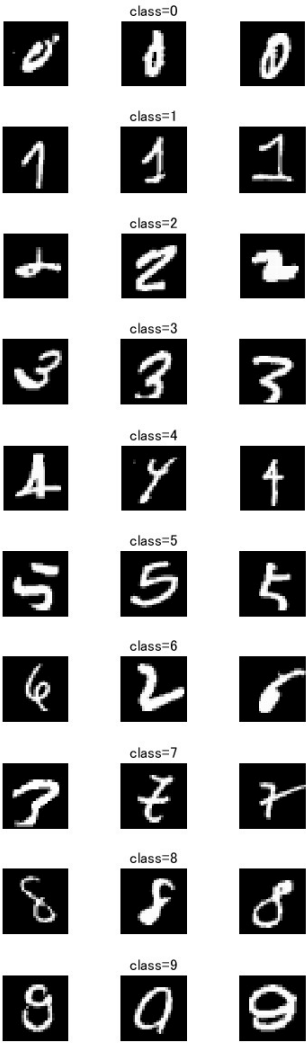


Fig.5 shows the 3 examples that were most frequently selected, across boosting iterations, as the example of largest weight.

Fig.6 show the visualized weak learners. A white pixel denotes a feature for regular weak learner, which outputs 1 for a value greater than its threshold. A black pixel represents a feature for a twin weak learner, which outputs -1 for a value greater than its threshold. Therefore, these images have vague shapes of the corresponding digits.

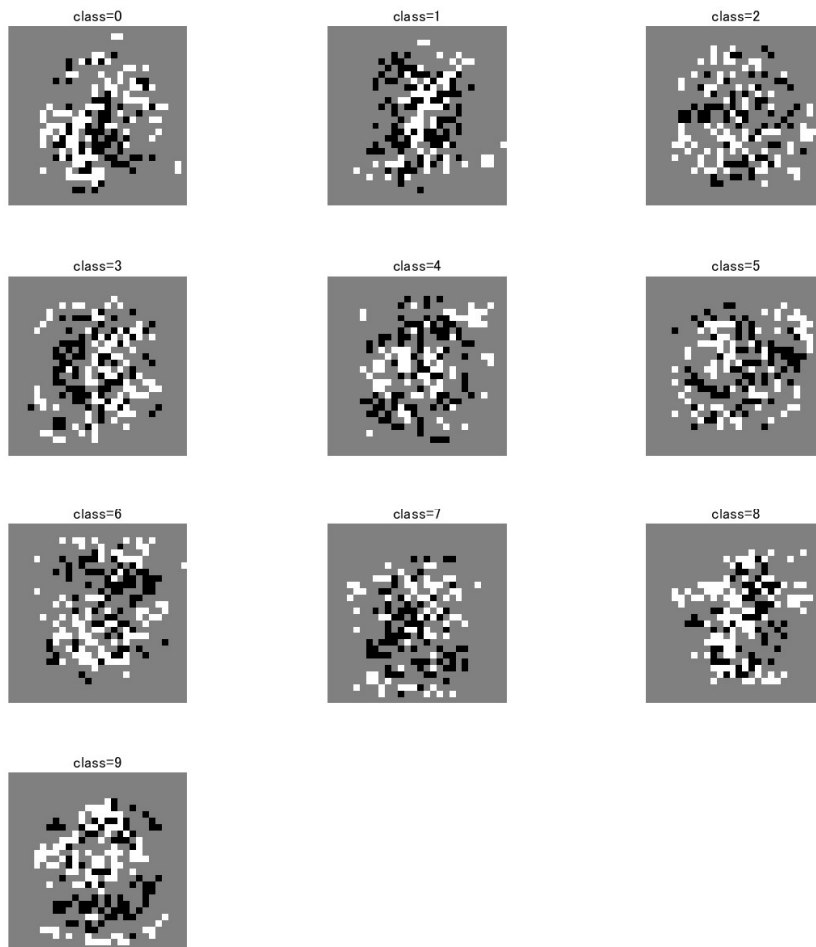


Fig.6 the visualized weak learners.