

Homework 8, due November 1st, 11:59pm

October 25, 2017

1. Implement Logitboost using 1D linear regressors as weak learners. At each boosting iteration choose the weak learner that obtains the largest reduction in the loss function on the training set $D = \{(\mathbf{x}_i, y_i), i = 1, \dots, N\}$:

$$L = \sum_{i=1}^N \ln(1 + \exp[-y_i h(\mathbf{x}_i)]) \quad (1)$$

where $h(\mathbf{x}) = h_1(\mathbf{x}) + \dots + h_k(\mathbf{x})$ is the boosted classifier and y_i take values ± 1 . Alternatively you can use a package that provides LogitBoost or GradientBoost with the Logistic loss above.

- a) Using the `Gisette` data, train a Logitboost classifier on the training set, with $k \in \{10, 30, 100, 300\}$ boosted iterations. Plot the training loss vs iteration number for $k = 300$. Report in a table the misclassification errors on the training and test set for the models obtained for all these k . Plot the misclassification error on the training and test set vs k . (6 points)
- b) On any of the other six **binary classification** datasets, using the training and test sets specified in the syllabus, repeat point a), adding the misclassification errors to the table from a). (1-3 points each as specified in the syllabus)