Homework 8, due October 30th, 11:59pm

October 17, 2019

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 $D = \{(\mathbf{x}_i, y_i), i = 1, ..., N\}$, with $y_i \in \{0, 1\}$:

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

where $\tilde{y}_i = 2y_i - 1$ take values ± 1 and $h(\mathbf{x}) = h_1(\mathbf{x}) + ... + h_k(\mathbf{x})$ is the boosted classifier. Please note that the Logitboost algorithm from the slides uses $y_i \in \{0, 1\}$ and the loss uses $\tilde{y} \in \{-1, 1\}$.

a) Using the `Gisette` data, train a Logitboost classifier on the training set, with $k \in \{10, 30, 100, 300, 500\}$ boosted iterations. Plot the training loss vs iteration number for $k = 500$. Report in a table the misclassification errors on the training and test set for the models obtained for all these k. Plot the misclassification errors on the training and test set vs k. (5 points)

b) Repeat point a) on the `dexter` dataset. (2 points)

c) Repeat point a) on the `madelon` dataset. (2 points)