

Homework 5, due October 13th, 11:59pm

October 5, 2023

1. Implement the FSA variable selection method for linear models and binary classification with the logistic loss, as described in page 4 of the FSA slides. Use the parameters $s = 0.0001$, $\mu = 200$, $N^{iter} = 300$. Take special care to **normalize each column** of the X matrix to have zero mean and variance 1 and to use the same mean and standard deviation that you used for normalizing the train set also for normalizing the test set.

- a) Using the `Gisette` data, train a FSA classifier on the training set, starting with $\beta^{(0)} = 0$ to select $k \in \{10, 30, 100, 300, 500\}$ features. Plot the training loss vs iteration number for $k = 30$. 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 . Also plot the train and test ROC curves of the obtained model with 100 features. (4 points)
- b) Repeat point a) on the `dexter` dataset. (2 points)
- c) Repeat point a) on the `madelon` dataset. (3 points)