Homework 3, due September 19th, 11:59pm

September 12, 2018

1. Implement the Logistic Regression learning by gradient ascent as described in class. Before using logistic regression, be sure to normalize the variables of the training set to have zero mean and standard deviation 1, and to do the exact same transformation to the test set, using the mean and standard deviation of the training set.

a) Using the `Gisette` data, train a logistic regressor on the training set, starting with $\mathbf{w}^{(0)} = 0$, with 300-1000 gradient ascent iterations and shrinkage $\lambda = 0.001$ in the update equation:

$$\mathbf{w}^{(t+1)} \leftarrow \mathbf{w}^{(t)} - \eta\lambda\mathbf{w}^{(t)} + \frac{\eta}{N}\frac{\partial}{\partial \mathbf{w}}L(\mathbf{w}^{(t)})$$

Observe that there is an extra factor of $1/N$ in the loss term compared to the class notes.

Find a good learning rate η such that the log-likelihood converges in 300-1000 iterations and is monotonically increasing. Plot the log-likelihood vs iteration number. Report in a table the misclassification error on the training and test set. (6 points)

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

c) Repeat point a) on the `hill-valley` dataset. For this data you might have to go up to 10,000 iterations for the algorithm to converge. (2 points)