Homework 8, due March 4th, 11:59pm

February 27, 2020

1. Download the file `gisette.zip` from Canvas. It contains training and validation files for the Gisette dataset, a binary classification dataset.

The training examples are $(\mathbf{x}_i, y_i), i = 1, ..., n$ where $\mathbf{x}_i$ are rows in the `gisette_train.data` file, and the y_i are the corresponding labels from `gisette_train.labels`.

The test examples are in `gisette_valid.data` file, and the corresponding labels in `gisette_valid.labels`.

We want to minimize the following loss function:

$$L(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n} \log(1 + \max(0, 1 - y_i \mathbf{x}_i^T \mathbf{w})^2) + \lambda \mathbf{w}^T \mathbf{w}, \text{ with } \lambda = 0.0001.$$

Before performing the minimization, be sure to normalize the $\mathbf{x}$ 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) Train the linear model on the training set, starting with $\mathbf{w}^{(0)} = 0$, with 300 gradient descent iterations. Find a good learning rate η such that the $L(\mathbf{w})$ converges in about 100-300 iterations and is monotonically decreasing. Plot the loss $L(\mathbf{w})$ vs iteration number. (2 points)

b) Repeat point a) using 31 iterations of the Newton method with $\gamma = 1$. (3 points)

c) Repeat point a) using 31 iterations of the BFGS algorithm. (2 points)

d) Repeat point a) using 31 iterations of the L-BFGS algorithm. (2 points)

e) Let's use the minimum value from all four losses obtained at a),b),c),d) as the target L_0. On the same figure, plot $\log(L(\mathbf{w}) - L_0)$ vs iteration number for the first 30 iterations for the four loss functions obtained above. This illustrates the convergence rate of the four algorithms. (1 point)