Stochastic gradient for support vector machine

(Due in class March-9-2018)

1 Problem description

The original linear support vector machine uses the hinge loss function $\max(0, 1-t)$ and is formulated as

$$\min_{\mathbf{w},b} \frac{1}{N}\sum_{i=1}^{N} \max\left(0, 1-y_i(b+\mathbf{x}_i^\top \mathbf{w})\right). \tag{1}$$

The hinge loss function is non-differentiable and causes trouble on designing efficient solvers. In our first assignment, we see that one way to conquer the trouble is to smooth the loss function by the huberized hinge loss function, which leads to the so-called huberized support vector machine (HSVM).

Similar to the elastic-net regularized HSVM, we can also use the regularization term in (1), i.e.,

$$\min_{\mathbf{w},b} \frac{1}{N}\sum_{i=1}^{N} \max\left(0, 1-y_i(b+\mathbf{x}_i^\top \mathbf{w})\right) + \lambda_1\|\mathbf{w}\|_1 + \frac{\lambda_2}{2}\|\mathbf{w}\|^2. \tag{2}$$

2 Requirements

Include every item below in a single report and attach your code. Use the larger-sized dataset that you used in the first assignment.

1. Develop a solver using the stochastic gradient for the elastic-net regularized HSVM, and compare it to the solver that you developed in the first assignment by using the accelerated proximal gradient. You can plot the objective values at the iterates generated by the two solvers and also report their prediction accuracy on the testing dataset.

2. Develop a solver using the stochastic gradient for (2) and compare its prediction accuracy to that given by the solver for the elastic-net regularized HSVM. You can vary the value of the smoothing parameter δ (but with the same λ_1 and λ_2) to see difference between the two models.

3. Implement the SGVR and SAGA on the elastic-net regularized HSVM. You can mute b by fixing it to zero and choose a positive λ_2. Compare their performance to the stochastic gradient and the accelerated proximal gradient by plotting the objective values in terms of epoch.