Homework Assignment 2
Loss Functions and Support Vector Machines

Lecturer: Kyunghyun Cho

February 8, 2018

1. Equivalence of negative log probability and logistic loss (10 points) After replacing the label set from $\{0, 1\}$ to $\{-1, 1\}$, we introduced the log loss

$$D_{\text{log}}(y, \mathbf{x}; M) = \frac{1}{\log 2} \log(1 + \exp(-s(y, \mathbf{x}; M))),$$

as an alternative to the logistic regression distance function above. Show that these two are equivalent up to a constant multiplication for logistic regression.

2. Hinge loss gradients (10 points) Unlike the log loss, the hinge loss, defined below, is not differentiable everywhere:

$$D_{\text{hinge}}(y, \mathbf{x}; M) = \max(0, 1 - s(y, \mathbf{x}; M)).$$

Does it mean that we cannot use a gradient-based optimization algorithm for finding a solution that minimizes the hinge loss? If not, what can we do about it?

3. Model Selection (20 points due to importance) Consider that we are learning a logistic regression $M^{(1)}$ and a perceptron $M^{(2)}$, and we have three dataset partitions: a training set D_{train}, a validation set D_{val}, and a test set D_{test}.

The two models are iteratively optimized on D_{train} over T steps, and now we have T logistic regression parameter configurations (i.e. weights and biases) $M_1^{(1)}, M_2^{(1)}, \ldots, M_T^{(1)}$ and T perceptron configurations $M_1^{(2)}, M_2^{(2)}, \ldots, M_T^{(2)}$, all with different parameters.

We now evaluate the expected cost for all the $2T$ models on training set, validation set, and test set. So we have $6T$ quantities $\tilde{R}^{(i)}_{\text{train},t}$, $\tilde{R}^{(i)}_{\text{val},t}$, $\tilde{R}^{(i)}_{\text{test},t}$ where $i = 1, 2$ and $t = 1, \ldots, T$.

1. Which i and t should we pick as the best model? (10 points)
2. How should we report the generalization error? (10 points)

4. Image Recovery & Numerical Stability (20 points) Programming Assignment: Please download