Alperen Karaoglu N13012011

Homework Assignment 2
Loss Functions and Support Vector Machines

Lecturer: Kyunghyun Cho
January 30, 2017

1. The logistical regression distance function above yields 1 for a score below 0 and 0 for a score above 0. The log loss function with labels set to {-1,1} instead of {0,1} is a monotonically decreasing function that yields above 1 when the score is below 0 and below 1 when the score is above 0 which means that the result of the classification will not change up to a constant multiplication.

2. Because the hinge loss function is non-differentiable everywhere we cannot use the normal gradient descent algorithm to minimize it. The solution to this problem is provided by using sub-gradient descent instead, we compute the sub-gradient by estimating data points for which hinge loss is greater than zero and then we apply it with a step size that is decreasing in time.

3. From the class notes we have

Then, the margin can be defined in terms of these two distances as

$$\gamma = \frac{1}{2}(d^+ + d^-) \tag{1.20}$$

$$= \frac{C}{\|\mathbf{w}\|}, \tag{1.21}$$

where C is the unnormalized distance to the positive and negative examples from the decision boundary. These two examples are equi-distance C away from the decision

We want to maximize the margin where ||w|| is always positive which means that $\frac{C}{||w||}$ is monotonically decreasing. So mathematically minimizing ||w|| is equivalent to minimizing the margin.