CS 419M Introduction to Machine Learning **Spring 2021-22**

Lecture 2: Loss Functions

Lecturer: Abir De *Scribe: Parin Senta*

2.1 What conditions/things can be relaxed so that the above optimisation problem can be solved by a normal computer?

We had formulated the above problem on the assumption that we have "Oracle", a computer that can solve any problem (for which a solution exists) within a millisecond. However that's not practically possible. Hence certain conditions need to be relaxed.

2.1.1 Restricting type of h(x)

So far, all types of h(x) were considered for optimisation i.e there was no restriction on the nature of h(x). However for simplification, we only consider linear h(x). For the time being we do not think about if this simplification provides a minimum solution. So, h(x) takes the form $W^T X$ and hence the optimisation problem reduces to

$$\min_W \sum I\left(\frac{1 + sign(W^T X)}{2} \neq y\right) \tag{2.1}$$

2.1.2 Replacing Indicator function by Modulus

It is difficult for a normal computer to optimise with the Indicator function, hence replacing it with modulus is another restriction imposed in the problem.

$$\min_W \sum \left|\frac{1 + sign(W^T X)}{2} - y\right| \tag{2.2}$$

2.1.3 Making the Loss Function Differentiable

In order for us to apply various calculus ideas, its better to have a differentiable function. Since sign(x) is not differentiable , we simply get rid of it. To take care of modulus, we square the terms of summation. In the end, we have the foloowing loss function:

$$\min_W \sum \left(\frac{1 + W^T X}{2} - y\right)^2 \tag{2.3}$$

2.1.4 Limiting the Bounds of $W^T X$

There is a problem with the previous restriction. Consider the case when y=1 and $W^T X = 10$. Since $W^T X$ is positive it should be classified as 1. However its contribution to the overall loss function is high.(should be zero for a properly chosen loss function). This is due to the fact that $W^T X$ is unbounded. To take care of this we give $W^T X$ as an input to sigmoid function which is bound between 0 and 1. Sigmoid function is defined by $S(x) = \frac{1}{1+e^-x}$.

$$\min_W \sum \left(S(W^T X) - y \right)^2 \tag{2.4}$$

2.2 Group Details and Individual Contribution