DSC 140A - Discussion 03

Problem 1.

A subgradient of the absolute loss is:

$$\begin{cases} \operatorname{Aug}(\vec{x}), & \text{if } \operatorname{Aug}(\vec{x}) \cdot \vec{w} - y > 0, \\ -\operatorname{Aug}(\vec{x}), & \text{if} \operatorname{Aug}(\vec{x}) \cdot \vec{w} - y < 0, \\ \vec{0}, & \text{otherwise.} \end{cases}$$

Suppose you are running subgradient descent to minimize the risk with respect to the absolute loss in order to train a function $H(x) = w_0 + w_1 x$ on the following data set:

x	y
1	3
2	5
3	7

Suppose that the initial weight vector is $\vec{w} = (0,0)^T$ and that the learning rate $\eta = 1$. What will be the weight vector after one iteration of subgradient descent?

Problem 2.

Consider the function $f(\vec{z}) = f(z_1, z_2) = \max(z_1, z_2)$.

a) Using the definition of the subgradient, check if $(1,1)^T$ is a subgradient at the point $(2,2)^T$.

b) Show that $(1,0)^T$ is a valid subgradient at $(2,2)^T$.

c) Show that $(\frac{1}{2}, \frac{1}{2})^T$ *is* a subgradient at $(2,2)^T$.

Problem 3.

The absolute loss of a linear predictor is

$$\ell_{\text{abs}}(\operatorname{Aug}(x) \cdot \vec{w}, y) = |\operatorname{Aug}(x) \cdot \vec{w} - y|.$$

We can write this as a piecewise function:

$$\ell_{\text{abs}}(\operatorname{Aug}(x) \cdot \vec{w}, y) = \begin{cases} \operatorname{Aug}(x) \cdot \vec{w} - y, & \text{if } \operatorname{Aug}(x) \cdot \vec{w} - y > 0, \\ y - \operatorname{Aug}(x) \cdot \vec{w}, & \text{if } \operatorname{Aug}(x) \cdot \vec{w} - y < 0, \\ 0, & \text{if } \operatorname{Aug}(x) \cdot \vec{w} = y. \end{cases}$$

This loss function is not differentiable at $\operatorname{Aug}(x) \cdot \vec{w} = y$, but has a well-defined gradient everywhere else.

a) What is the gradient of the absolute loss with respect to $\vec{w}$ when $\operatorname{Aug}(x) \cdot \vec{w} - y > 0$?

b) What is the gradient of the absolute loss with respect to $\vec{w}$ when $\operatorname{Aug}(x) \cdot \vec{w} - y < 0$?

c) **Optional**: Show that $\vec{0}$ is a subgradient of the absolute loss at $\operatorname{Aug}(x) \cdot \vec{w} = y$.

Problem 4.

Using the definition, show that the function $f(\vec{w}) = a\vec{x} \cdot \vec{w} - b$ is convex as a function of $\vec{w}$, where $a, b \in \mathbb{R}$ and $\vec{x}, \vec{w} \in \mathbb{R}^d$.

Problem 5.

Let $f_1(\vec{w})$ and $f_2(\vec{w})$ be convex functions from $\mathbb{R}^d$ to $\mathbb{R}$. Define

$$f(\vec{w}) = \max\{f_1(\vec{w}), f_2(\vec{w})\}.$$

Show that $f(\vec{w})$ is convex.

Problem 6.

Recall that the Perceptron loss is:

$$L_{\text{perc}}(\vec{w}, \vec{x}, y) = \begin{cases} 0, & \text{if } \operatorname{sign}(\vec{w} \cdot \vec{x}) = y \text{ (correctly classified)}, \\ |\vec{w} \cdot \vec{x}|, & \text{if } \operatorname{sign}(\vec{w} \cdot \vec{x}) \neq y \text{ (misclassified)}. \end{cases}$$

Using the trick that $-y\,\vec{w} \cdot \vec{x} = |\vec{w} \cdot \vec{x}|$ in the case of misclassification, this can be be written in the equivalent form:

$$L_{\text{perc}}(\vec{w}, \vec{x}, y) = \max\{0, -y\,\vec{w} \cdot \vec{x}\}$$

Argue that the perceptron loss is convex as a function of $\vec{w}$.