DSC 140A - Midterm 01 Review

Problem 1.

Suppose $f(\vec{x})$ is a **convex** function of $\vec{x}$ and that the gradient of f at the point $\vec{x}^{(1)} = (0,5)^T$ is $(-3,5)^T$.

Let $\vec{x}^* = (x_1^*, x_2^*)$ be the minimizer of f; you can assume that $\vec{x}^* \neq \vec{x}^{(1)}$, which implies that $f(\vec{x}^*) < f(\vec{x}^{(1)})$. True or False: it must be the case that $x_2^* \geq 5$.

Solution:

False as the gradient is the direction of maximal ascent. We would move in the direction of $(3,-5)^T$ and therefore $x_2^* \leq 5$.

Problem 2.

Recall that a subgradient of the absolute loss is:

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

Suppose you are running subgradient descent to minimize the risk with respect to the absolute loss in order to train a function $H(x) = w_0 + w_1 x_1 + w_2 x_2$ on the following data set:

x_1	x_2	y
2	5	6
1	2	3

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

Solution:

As $\text{Aug}(\vec{x}) \cdot \vec{w} = 0$, $\text{Aug}(\vec{x}) \cdot \vec{w} - y = -y < 0$. So, $\frac{\partial}{\partial \vec{w}} R(\vec{w}) = \frac{1}{n} \Sigma_{i=1}^n \frac{\partial}{\partial \vec{w}} L_{abs}$. Plugging in our data we have, $\frac{\partial}{\partial \vec{w}} R(\vec{w}) = -\frac{1}{2}\left(\begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \right) = -\frac{1}{2} \begin{bmatrix} 2 \\ 3 \\ 7 \end{bmatrix}$ After 1 iteration of gradient descent $\vec{w} = (0,0,0)^T + \frac{1}{2}(2,3,7)^T = (1, 3/2, 7/2)^T$.

Problem 3.

Let $X = \{\vec{x}^{(i)}, y_i\}$ be a data set of n points where each $\vec{x}^{(i)} \in \mathbb{R}^d$.

Let $Z = \{\vec{z}^{(i)}, y_i\}$ be the data set obtained from the original by standardizing each feature. That is, if a matrix were created with the i-th row being $\vec{z}^{(i)}$, then the mean of each column would be 0, and the variance would be 1.

a) Suppose linear predictors H_1 and H_2 are fit on X and Z by minimizing the MSE, respectively.

True or False: $H_1(\vec{x}^{(i)}) = H_2(\vec{z}^{(i)})$ for every $i = 1, \ldots, n$.

Solution:

True. In linear functions, changing input unit do not change the output.

b) Suppose that X and Z are both linearly-separable. Suppose Hard-SVMs H_1 and H_2 are trained on X and Z, respectively.

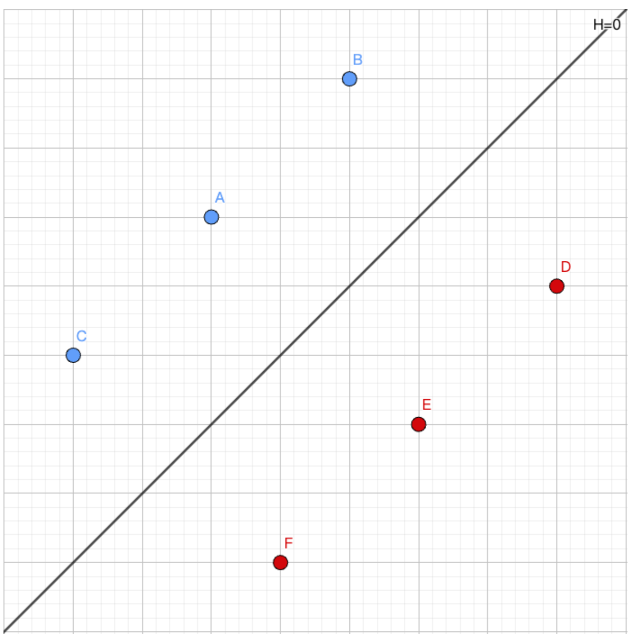
True or False: $H_1(\vec{x}^{(i)}) = H_2(\vec{z}^{(i)})$ for every $i = 1, \ldots, n$.

Solution:

False. In (Hard) SVMs, changing scales may change the output, e.g., by changing the set of support vectors.

Problem 4.

Consider the image below:



The blue points have label $+1$, and the red points have label -1. Suppose H is a linear prediction function, and when H is applied to the point A in the above image, $H(\vec{x}) = 4$. The black line in the middle of the image is the decision boundary $H = 0$.

You may assume that H is exactly in the middle of the points, and that all points are equidistant from the decision boundary.

a) What is the mean square loss of this prediction function, H?

Solution:

The output of H on the blue points is 4, while for the red points it is -4. This follows from the fact that all points are equidistant from the decision boundary, and H is linear. The target label for the blue and red points is 1 and -1, respectively. So the prediction is off by 3 in each case. This makes the squared error on each point $3^2 = 9$, so the mean squared error (mean square loss) is 9.

b) True or false: there exists a linear prediction function H which has a mean square loss of zero on this data.

> **Solution:**
>
> True. One simple way is to modify H. Let $H'(\vec{x}) = H(\vec{x})/4$, then H' is also a linear prediction function with its output exactly correct for each data point.
>
> Alternatively, here's a geometric argument: a linear prediction function in two dimensions fits a plane to the data, where we have lifted to blue points to a height of 1 and the red points to a height of -1. Because the points are all equidistant from the black line in the figure above, there is a way to draw a plane that goes exactly through all of the points.

Problem 5.

a) True or false: Assume $f(x) = f_1(x) + f_2(x) + \ldots + f_k(x)$. If f is convex, then all $f_1, \ldots, f_k$ are convex.

> **Solution:**
>
> False. (The inverse of the property that the sum of convex functions is also convex is false). Consider the counter-example where $f(x) = 0$ and $f_1(x) = -f_2(x)$.

b) True or false: Assume $f(x) = g(h(x))$. If f is convex, then both g and h are convex.

> **Solution:**
>
> False. (The inverse of the property that the composition of convex functions is also convex is false). Consider the counter-example where $f(x) = x(x \geq 0)$ and $g(x) = x^2, h(x) = \sqrt{x}$, where h is non-convex.

c) True or false: If f is convex, then $-f$ is non-convex.

> **Solution:**
>
> False. Consider the counter-example $f(x) = 0$.
>
> Note that: If f is convex, then $-f$ is concave.