University of Waterloo
CS 480
r2knowle: 2023-10-22

Exercise # 2

Q1) To begin we are given that for any $x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$ that the error term is:

$$C\sum_{i=1}^{n} \max\{|y_i - (w^T x_i + b)| - \epsilon, 0\}$$

We will then introduce a term y_i such that $y_i >= \max\{|y_i - (w^T x_i + b)| - \epsilon, 0\}$. Plugging this into our original equation we given in the question we get:

$$\max_{\alpha,\beta} \min_{w,b} \frac{1}{2}||w||_2^2 + \sum_{i=1}^{n}(a_i y_i + a_i(|y_i - (w^T x_i + b)| - \epsilon) - \beta y_i)$$

Before we take the derivative in respect to b, and w. We will start by considering both sides of the absolute value. Starting with $y_i - (w^T x_i + b) - \epsilon > 0$ we get:

$$\frac{d}{db} = -\sum_{i=1}^{n} a_i = 0$$

$$\frac{d}{dw} = w - \sum_{i=1}^{n} a_i x_i = 0$$

$$\frac{d}{dy_i} = C + a_i + \beta_i = 0$$

Now since $w = \sum_{i=1}^{n} a_i x_i$ we can get $||w||_2^2 = \sum_{i=0}^{n}\sum_{j=1}^{n} a_i a_j < x_i, x_j >$ plugging this back in to our equation gives us get:

$$\begin{aligned}
&= \max_{\alpha,\beta} \min_{w,b} \frac{1}{2}||w||_2^2 + \sum_{i=1}^{n} a_i(y_i - (w^T x_i + b)) \\
&= \max_{\alpha,\beta} \frac{1}{2}\sum_{i=0}^{n}\sum_{j=1}^{n} a_i a_j < x_i, x_j > + \sum_{i=1}^{n} a_i(y_i - (w^T x_i + b)) \\
&= \max_{\alpha,\beta} \frac{1}{2}\sum_{i=0}^{n}\sum_{j=1}^{n} a_i a_j < x_i, x_j > + \sum_{i=1}^{n} a_i y_i - \sum_{j=1}^{n}\sum_{i=1}^{n} a_i a_j < x_i, x_j > - \sum_{i=1}^{n} b\alpha_i \\
&= \max_{\alpha,\beta} \sum_{i=1}^{n} a_i y_i - \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} a_i a_j < x_i, x_j > - \sum_{i=1}^{n} b\alpha_i \\
&= \max_{\alpha,\beta} \sum_{i=1}^{n} a_i y_i - \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} a_i a_j < x_i, x_j > \;\; s.t\; C = \alpha + \beta, \sum_{i=1}^{n} a_i = 0
\end{aligned}$$

On the other hand if $-y_i + (w^T x_i + b) + \epsilon > 0$:

$$\frac{d}{db} = \sum_{i=1}^{n} a_i = 0$$

$$\frac{d}{dw} = w + \sum_{i=1}^{n} a_i x_i = 0$$

$$\frac{d}{dy_i} = C + a_i + \beta_i = 0$$

Now since $w = -\sum_{i=1}^{n} a_i x_i$ we can get $||w||_2^2 = \sum_{i=0}^{n} \sum_{j=1}^{n} a_i a_j < x_i, x_j >$ plugging this back in to our equation gives us get:

$$\begin{aligned}
&= \max_{\alpha,\beta} \min_{w,b} \frac{1}{2}||w||_2^2 + \sum_{i=1}^{n}(a_i(-y_i + (w^T x_i + b)) \\
&= \max_{\alpha,\beta} \frac{1}{2} \sum_{i=0}^{n} \sum_{j=1}^{n} a_i a_j < x_i, x_j > + \sum_{i=1}^{n} a_i(-y_i + (w^T x_i + b))) \\
&= \max_{\alpha,\beta} \frac{1}{2} \sum_{i=0}^{n} \sum_{j=1}^{n} a_i a_j < x_i, x_j > - \sum_{i=1}^{n} a_i y_i - \sum_{j=1}^{n} \sum_{i=1}^{n} a_i a_j < x_i, x_j > + \sum_{i=1}^{n} b\alpha_i \\
&= \max_{\alpha,\beta} - \sum_{i=1}^{n} a_i y_i - \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} a_i a_j < x_i, x_j > + \sum_{i=1}^{n} b\alpha_i \\
&= \max_{\alpha,\beta} - \sum_{i=1}^{n} a_i y_i - \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} a_i a_j < x_i, x_j > \;\; s.t\ C = \alpha + \beta, \sum_{i=1}^{n} a_i = 0
\end{aligned}$$

Thus giving us the Lagrangian Dual for both sides of the absolute value.

Q2) To begin we know we are given:

$$C\sum_{i=1}^{n}\max\{|y_i-(w^Tx_i+b)|-\epsilon,0\}$$

this can split the max equation into two separate forms, either its 0 or:

$$|y_i-(w^Tx_i+b)|-e$$

This in and of itself has two forms its either $y_i-(w^Tx_i+b)-e$ or $-y_i+(w^Tx_i+b)-e$. Now we will state that:

$$y_i-(w^Tx_i+b)-e<-y_i+(w^Tx_i+b)-e$$

It then would follow that $y_i-(w^Tx_i+b)$ would be less then 0, and thus we only need to include $-y_i+(w^Tx_i+b)$ when looking at the gradient. Thus if we take the gradients we get:

$$\frac{d}{db}=C$$

$$\frac{d}{dw}=C\sum_{i=1}^{n}x_i$$

In the cases where $-y_i+(w^Tx_i+b)$ would be less then 0, and thus we only need to include $y_i-(w^Tx_i+b)$ when looking at the gradient. Thus if we take the gradients we get:

$$\frac{d}{db}=-C$$

$$\frac{d}{dw}=-C\sum_{i=1}^{n}x_i$$

And so if we combine all three we get that the sub gradient for w will be:

$$\frac{d}{dw}=\begin{cases}-C\sum_{i=1}^{n}x_i & y_i-(w^Tx_i+b)>\epsilon\\ 0 & |y_i-(w^Tx_i+b)|<\epsilon\\ C\sum_{i=1}^{n}x_i & -y_i+(w^Tx_i+b)>\epsilon\end{cases}$$

And for b we get:

$$\frac{d}{db}=\begin{cases}-nC & y_i-(w^Tx_i+b)>\epsilon\\ 0 & |y_i-(w^Tx_i+b)|<\epsilon\\ nC & -y_i+(w^Tx_i+b)>\epsilon\end{cases}$$

Q3) We are given the equation:

$$p^{\eta}(w)=\min_{z}\frac{1}{2\eta}||z-w||_2^2+\frac{1}{2}||z||_2^2$$

This will be minimized when z is equal to the previous iteration for w. in other words we will have that:

$$w^(t+1)=w-\eta z$$

Q3) After running gradient descent (included as python file) we get the following training error, training loss and test error:

$$\text{Training Error} = 0.7702098196531493$$

$$\text{Test Error} = 0.5621596024609502$$

$$\text{Loss} = 0.8432394036914252$$