CS221 Problem Workout

Week 1

1) Problem 1: Gradient and Gradient Descent

(i) Let $\phi(x) : \mathbb{R} \mapsto \mathbb{R}^d$, $w \in \mathbb{R}^d$. Consider the following objective function (a.k.a. loss function).

$$\text{Loss}(x, y, w) = \begin{cases} 
1 - 2(w \cdot \phi(x))y & \text{if } (w \cdot \phi(x))y \leq 0 \\
(1 - (w \cdot \phi(x))y)^2 & \text{if } 0 < (w \cdot \phi(x))y \leq 1 \\
0 & \text{if } (w \cdot \phi(x))y > 1,
\end{cases}$$

where $y \in \mathbb{R}$. Compute the gradient $\nabla_w \text{Loss}(x, y, w)$.

(ii) Write out the Gradient Descent update rule for some function $\text{TrainLoss}(w) : \mathbb{R}^d \mapsto \mathbb{R}$. 


(iii) Let $d = 2$ and $\phi(x) = [1, x]$. Consider the following loss function.

$$
\text{TrainLoss}(w) = \frac{1}{2} \left( \text{Loss}(x_1, y_1, w) + \text{Loss}(x_2, y_2, w) \right).
$$

(1)

Compute $\nabla_w \text{TrainLoss}(w)$ for the following values of $x_1, y_1, x_2, y_2, w$.

$$
w = \begin{bmatrix} 0, \frac{1}{2} \end{bmatrix},
$$

$$
x_1 = -2, \quad y_1 = 1,
$$

$$
x_2 = -1, \quad y_2 = -1.
$$

(iv) Perform two iterations of Gradient Descent to minimize the objective function $\text{TrainLoss}(w) = \frac{1}{2} \left( \text{Loss}(x_1, y_1, w) + \text{Loss}(x_2, y_2, w) \right)$ with values for $x_1, y_1, x_2, y_2$ as above. Use initialization $w^0 = [0, \frac{1}{2}]$ and step size $\eta = \frac{1}{2}$. 

2) Problem 2: Gradient computation

(i) Let $\phi(x) : \mathbb{R} \mapsto \mathbb{R}^d$, $w \in \mathbb{R}^d$, and $f(x, w) = w \cdot \phi(x)$. Consider the following loss function.

$$\text{Loss}(x, y, w) = \frac{1}{2} \max\{2 - (w \cdot \phi(x))y, 0\}^2.$$  

(2)

Compute its gradient $\nabla_w \text{Loss}(x, y, w)$. 

3) **Problem 3: Vector visualization**

Recall that we can visualize a vector \( \mathbf{w} \in \mathbb{R}^d \) as a point in \( d \)-dimensional space. Let us now visualize some vectors in 2 dimensions on pen and paper.

(i) Consider \( \mathbf{x} \in \mathbb{R}^2 \). Draw the line (i.e. the “decision boundary”) that separates between vectors having a positive dot product with weights \( \mathbf{w} = [3, -2] \) and those having a negative dot product. Shade the part of the 2D plane that contains vectors satisfying \( \mathbf{w} \cdot \mathbf{x} > 0 \).

Hint: It might help to write out the expression for the dot product and seeing the relation between \( x_1 \) and \( x_2 \) that leads to a positive dot product. You could also use the geometric interpretation of the dot product.
(ii) Repeat the above for $\mathbf{w} = [2, 0]$ and $\mathbf{w} = [0, 2]$. 
(iii) A small twist: visualize the set of vectors where \( \mathbf{w} \cdot \mathbf{x} \geq 1 \) for \( \mathbf{w} = [3, -2] \).
(iv) Consider the following element-wise inequality notation. For two vectors \( \mathbf{a}, \mathbf{b} \in \mathbb{R}^d \),

\[
\mathbf{a} \leq \mathbf{b} \iff a_i \leq b_i \ \forall i = 1, 2, \ldots d.
\]  

(3) Suppose we have a matrix \( \mathbf{A} \in \mathbb{R}^{2 \times 2} \) and a vector \( \mathbf{b} \in \mathbb{R}^2 \) as follows.

\[
\mathbf{A} = \begin{bmatrix} 3 & -2 \\ 2 & 0 \end{bmatrix}, \mathbf{b} = [1, 0].
\]  

(4) Visualize the set of vectors where \( \mathbf{A}\mathbf{x} \geq \mathbf{b} \). Hint: A matrix vector product is a collection of dot products, and the above set can be obtained by the intersection of two of the sets constructed in the previous questions.
4) Problem 4: More gradient computations

(i) Compute the gradient of the loss function below.

\[ \text{Loss}(x, y, w) = \sigma(- (w \cdot \phi(x)) y), \]

where \( \sigma(z) = (1 + \exp(-z))^{-1} \) is the logistic function.

(ii) Suppose we have the following loss function.

\[ \text{Loss}(x, y, w) = \max\{1 - [(w \cdot \phi(x)) y], 0\}, \]

where \( \lfloor a \rfloor \) returns \( a \) rounded down to the nearest integer. Determine what the gradient of this function looks like, and whether gradient descent is suitable to optimize this loss function.