CS221 Problem Workout

Week 1

Welcome to CS 221! During the problem sessions, we will work through practice homework and exam problems. If you have any questions, comments, concerns, feel free to reach out to the problem session lead Samantha Liu, and co-presenter Michael Ryan. Here are some key takeaways from this week:

Key Takeaways from this Week

1. **Syllabus**: Please take some time and review the course syllabus. If you have any questions or concerns, please reach out to the instructors and or CAs.
2. **Gradient and Gradient Descent**:
   - **Gradient**: the vector representing the direction and rate of the fastest increase of a scalar-valued function f at a point. ∇f can be derived by taking all partial derivatives of f and forming them into a vector-valued function.
   - **Gradient Descent**: an optimization algorithm for finding the local minimum of a differential function f by taking repeated steps in the opposite direction of the gradient at the current point.
3. **Loss Functions (Preview)**: The measure of how far the model's prediction is from the real (or ground truth) y.

1 Practice Problems

1) **Problem 1: Gradient computation**

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

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

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

2) Problem 2: More gradient computations

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

$$\text{Loss}(x, y, \mathbf{w}) = \sigma(-(\mathbf{w} \cdot \phi(x))y), \tag{2}$$

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

(ii) Suppose we have the following loss function.

$$\text{Loss}(x, y, \mathbf{w}) = \max\{1 - \lfloor(\mathbf{w} \cdot \phi(x))y\rfloor, 0\}, \tag{3}$$

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.

3) Problem 3: Gradient and Gradient Descent

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

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

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

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

$$\text{TrainLoss}(\mathbf{w}) = \frac{1}{2}\Big(\text{Loss}(x_1, y_1, \mathbf{w}) + \text{Loss}(x_2, y_2, \mathbf{w})\Big). \tag{4}$$

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

$$\begin{aligned} \mathbf{w} &= \left[0, \frac{1}{2}\right], \\ x_1 = -2,&\ \ y_1 = 1, \\ x_2 = -1,&\ \ y_2 = -1. \end{aligned}$$

(iii) Now, let's define the Gradient Descent update rule for some function $\text{TrainLoss}(\mathbf{w}) : \mathbb{R}^d \mapsto \mathbb{R}$. The rule helps us update the weights $\mathbf{w}$.

$$\mathbf{w} := \mathbf{w} - \eta \nabla_{\mathbf{w}} \text{TrainLoss}(\mathbf{w}), \text{ where } \eta \text{ is the step size.} \tag{5}$$

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

4) **Problem 4 (Extra): 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 $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$,

$$\mathbf{a} \leq \mathbf{b} \iff a_i \leq b_i \ \forall i = 1, 2, \ldots d. \tag{6}$$

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

$$A = \begin{bmatrix} 3 & -2 \\ 2 & 0 \end{bmatrix}, \mathbf{b} = [1, 0]. \tag{7}$$

Visualize the set of vectors where $A\mathbf{x} \geq \mathbf{b}$. 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.