CSE 5819 Assignment #2

by: Aayushi Verma (uef24001)

This assignment is organized as follows:

1 Part 1: ChatGPT Self-Learning (20pts)

You can start from the following prompts, but you need to create subsequent questions that attempt to understand the basic concepts of machine learning.

1.1 What defines a probability function?

1.2 What is Bernoulli distribution?

1.3 What is binomial distribution?

1.4 What is the relationship between Bernoulli and binomial distributions?

1.5 Properties of the expectation of a random variable X, E[X].

1.6 Is there a difference between Gaussian distribution and Normal distribution?

1.7 What is the density function of a Gaussian distribution?

1.8 What is the p-norm of a vector?

1.9 What is the formula of p-norm when $p = 0$ (or $p=1$, $p=\infty$)?

1.10 Why is any vector norm a convex function?

1.11 What is the definition of a convex function?

1.12 What is the definition of a convex set?

1.13 What is the definition of convex optimization?

2 Problems

2.1 title

1. [**Basic Statistics**] (15 pts) If $X \sim N(\mu, \sigma^2)$, $E(X) = \mu$, $Var[X] = \sigma^2$, and $E[X^2] = \mu^2 + \sigma^2$. Further, recall that expectation is linear, so *it* obeys the following three properties:

   $E[X + c] = E[X] + c$ for any constant c,
   $E[X + Y] = E[X] + E[Y]$,
   $E[aX] = aE[X]$ for any constant a.

   We note that if X and X' are independent, then $E[XX'] = E[X]E[X']$.
   Consider two points (sampled independently) from the same class follow: $X \sim N(\mu_1, \sigma^2)$ and $X' \sim N(\mu_1, \sigma^2)$. What is the expected squared distance between them, i.e., $E[(X - X')^2]$?

2.2 [Basic Linear Algebra] (15 pts) We are given a vector $x = [0, 0.2, 1.0, 2.2]$. Which of the following vector is closest to x and what is the distance from the closest point to x under each of the following vector norms?

$$x_1 = [0.7, 0.2, 0.5, 2.0]$$
$$x_2 = [0, 1.0, 1.5, 2.2]$$
$$x_3 = [0.8, 0.1, 1.2, 2.0]$$

a) 0-norm = b) 1-norm = c) 2-norm = d) ∞-norm =

2.3 title

3. [**Convexity**] (1) (10 pts) Given a vector space V, any vector norm satisfies the following three properties:

   (N1) $\|x\| \geq 0$, and $\|x\| = 0$ iff $x = 0$. **Non-negativity**
   (N2) $\|\lambda x\| = |\lambda| \, \|x\|$. **Scaling**
   (N3) $\|x + y\| \leq \|x\| + \|y\|$. **Triangle Inequality**

   If we consider this vector norm as a function of x, $f(x) = ||x||$, prove that $f(x)$ is a convex function (using the above properties). Recall the definition of a convex function is:
   A real-valued function $f : \mathbb{R}^n \to \mathbb{R}$ is called convex if, for any two points $x, y \in \mathbb{R}^n$ and any $\lambda \in [0, 1]$, the following inequality holds:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

2.4 (2) (10 pts). Prove that the convexity is preserved under a linear transformation. Supposed $f(w)$ is convex in terms of w. Prove that $g(w) = f(Xw + b)$ is also convex in terms of w where X is a fixed matrix of appropriate size, and b is a fixed vector of appropriate size. (So X and b are not variables in g). (Hint: you can simply use the definition of convex function)

2.5 (3) (10 pts). From (1) and (2), prove the "loss" function $||y - Xw||_2$ is convex in terms of w where $X_{n\times d}$ is a data matrix containing n examples and each example has d features, and y is a column vector of length d. This dataset has been observed and thus fixed, and w is the only variable in the norm.