Assigned: 26 August

Homework #1

EE 541: Fall 2024

Due: Friday, 06 September at 22:00. Submission instructions will follow separately on brightspace.

Use only Python standard library modules (https://docs.python.org/3/library/) and `matplotlib` for this assignment, *i.e.*, do not `import numpy`, `scikit`, or any other non-standard package.

1. An MLP has two input nodes, one hidden layer, and two outputs. The two sets of weights and biases are given by:

$$W_1 = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix} \qquad b_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 2 & 2 \\ 2 & -3 \end{bmatrix} \qquad b_2 = \begin{bmatrix} 0 \\ -4 \end{bmatrix}$$

The non-linear activation for the hidden layer is ReLU (rectified linear unit) – that is $h(x) = \max(x, 0)$. The output layer is linear (*i.e.*, identity activation function). The output for layer l is given by $a^{(l)} = h_l\left(W_l a^{(l-1)} + b_l\right)$. What is the output activation for input $x = [+1; -1]^T$?

2. Let $f(x, y) = 4x^2 + y^2 - xy - 13x$

   (a) Find $\frac{\partial f}{\partial x}$, the partial derivative of f with respect to x. Find $\frac{\partial f}{\partial y}$.

   (b) Find $(x, y) \in \mathbb{R}^2$ that minimizes f.

3. A hyper-plane in $\mathbb{R}^n$ is the set, $\left\{\mathbf{x} : \mathbf{x} \in \mathbb{R}^n, w^T x + b = 0\right\}$, where $w \in \mathbb{R}^n$ and b is a real scalar.

   (a) The solution of the following optimization problem describes the distance between a point $x_0 \in \mathbb{R}^n$ and the hyperplane $w^T x + b = 0$:

$$\min_x \|x_0 - x\|_2 \quad \text{s.t.} \quad w^T x + b = 0.$$

   Derive an analytic solution for the distance between x_0 and $w^T x + b = 0$.

   (b) What is the distance between two hyperplanes, $w^T x + b_1 = 0$ and $w^T x + b_2 = 0$?

4. A function $f(x)$ is convex if

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

for all x, y and $0 < \lambda < 1$.

   (a) Use this definition to prove that $f(x) = x^2$ is a convex function. Verify that $f(x) = x^3$ is not a convex function.

(b) An $n \times n$ matrix A is a positive semi-definite matrix if $\mathbf{x}^T A\mathbf{x} \geq 0$, for any $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{x} \neq \mathbf{0}$. Prove that the function $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is convex if A is a positive semi-definite matrix for $\mathbf{x} \in \mathbb{R}^n$.

5. Simulate tossing a biased coin (a Bernoulli trial) where $P[\text{HEAD}] = 0.70$.

   (a) Count the number of heads in 50 trials. Record the longest run of heads.

   (b) Repeat the 50-flip experiment 20, 100, 200, and 1000 times. Use matplotlib to generate a histogram showing the observed number of heads for each case. Comment on the limit of the histogram.

   (c) Simulate tossing the coin 500 times. Generate a histogram showing the heads run lengths.

6. Define the random variable $N = \min\{n : \sum_{i=1}^{n} X_i > 4\}$ as the smallest number of standard uniform random samples whose sum exceeds four. Generate a histogram using 100, 1000, and 10000 realizations of N. Comment on the expected value $E[N]$.