Fall 2022 CS165B: Introduction to Machine Learning – Homework 1

Due: **Sunday, Oct 16th, 11:59 pm PST**

1. (5 points) Let

$$A = \begin{bmatrix} 4 & 1 & 3 & 6 \\ 2 & 7 & 5 & 3 \end{bmatrix}, B = \begin{bmatrix} 0 & 4 \\ 7 & 6 \\ 5 & 8 \\ 3 & 11 \end{bmatrix}, C = \begin{bmatrix} -13 & 0 & 2 \\ 5 & 2 & 10 \\ 0 & 7 & 9 \end{bmatrix}, D = \begin{bmatrix} 5 & -3 & -7 \\ 4 & 0 & 10 \\ 7 & 3 & 11 \end{bmatrix}, E = \begin{bmatrix} -4 & 5 \\ 12 & 7 \end{bmatrix}.$$

If possible, compute the following:

(a) $(3B)^T$

(b) $(A - B)^T$

(c) $(2B^T - A)^T$

(d) $(C + 2D^T + E)^T$

(e) $(-A)^T E$

Note: T subscript means transpose.

2. (4 points) Let

$$A = \begin{bmatrix} 2 & 7 & 3 \\ 1 & 0 & 9 \\ -1 & 2 & 10 \end{bmatrix}, B = \begin{bmatrix} -2 & 0 & 3 \\ 2 & -1 & 7 \\ 6 & 4 & -3 \end{bmatrix}.$$

Is $AB = BA$? Justify your answer.

3. (8 points) Compute the $\ell_1/\ell_2/\ell_\infty$ norms of the following vectors:

- $[0, 0, 0]$
- $[1, 2, 3]$
- $[2, 4, 6]$. How are the norms of $[2, 4, 6]$ related to those of $[1, 2, 3]$?
- Can you find a vector with negative norms, why?

4. (4 points) Given $X = [x_1, x_2, \cdots, x_n] \in \mathbb{R}^{m\times n}$ where $x_i \in \mathbb{R}^m$ for all i, and $Y^T = [y_1, y_2, \cdots, y_n] \in \mathbb{R}^{p\times n}$ where $y_i \in \mathbb{R}^p$ for all i. Show that

$$XY = \sum_{i=1}^{n} x_i (y_i)^T.$$

5. (6 points) Given $X \in \mathbb{R}^{m\times n}$, show that the matrix $X^T X$ is symmetric and positive semi-definite. When is it positive definite?

6. (4 points) Given $g(x, y) = e^{(x+y)} + e^{3xy} + e^{y^4}$, compute $\frac{\partial g}{\partial x}$ and $\frac{\partial g}{\partial y}$.

7. (12 points) Consider the matrix

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 1 & 2 \\ 3 & 2 & 5 \end{pmatrix},$$

(a) Compute the eigenvalues and corresponding eigenvectors of A. You are allowed to use PYTHON to compute the eigenvectors (but not the eigenvalues). Please include the code that you used for computing eigenvectors.

(b) What is the eigen-decomposition of A?

(c) What is the rank of A?

(d) Is A positive definite?

(e) Is A positive semi-definite?

(f) Is A singular?

8. (12 points) Consider the perceptron in two dimensions: $h(\mathbf{x}) = \text{sign}\left(\mathbf{w}^T\mathbf{x}\right)$ where $\mathbf{w} = [w_0, w_1, w_2]^T$ and $\mathbf{x} = [1, x_1, x_2]^T$. Technically, $\mathbf{x}$ has three coordinates, but we call this perceptron two-dimensional because the first coordinate is fixed at 1.

   (a) (6 points) Show that the regions on the plane where $h(\mathbf{x}) = +1$ and $h(\mathbf{x}) = -1$ are separated by a line. If we express this line by the equation $x_2 = ax_1 + b$, what are the slope a and intercept b in terms of w_0, w_1, w_2?

   (b) (6 points) Draw a picture for the cases $\mathbf{w} = [1, 2, 3]^T$ and $\mathbf{w} = -[1, 2, 3]^T$.

   In more than two dimensions, the $+1$ and -1 regions are separated by a *hyperplane*, the generalization of a line.

9. (20 points) In this problem, we explore the perceptron learning algorithm further with data sets of different sizes and dimensions. Please include your code (.py or .ipynb) in your submission, and include your plots and explanations in the PDF you submit.

   (a) (5 points) Generate a linearly separable data set of size 20. Plot the examples $\{(\mathbf{x}_n, y_n)\}$ as well as the target function f on a plane. Be sure to mark the examples from different classes differently, and add labels to the axes of the plot.

   (b) (5 points) Run the perceptron learning algorithm on the data set above. Report the number of updates that the algorithm takes before converging. Plot the examples $\{(\mathbf{x}_n, y_n)\}$, the target function f, and the final hypothesis g in the same figure. Comment on whether f is close to g.

   (c) (3 points) Repeat everything in (b) with another randomly generated data set of size 20. Compare your results with (b).

   (d) (3 points) Repeat everything in (b) with another randomly generated data set of size 100. Compare your results with (b).

   (e) (4 points) Repeat everything in (b) with another randomly generated data set of size $1,000$. Compare your results with (b).

10. (15 points) Suppose that we have four colored boxes r (red), b (blue), g (green), and y (yellow). Box r contains 5 apples, 2 oranges, and 3 mangoes, box b contains 3 apple, 1 orange, and 7 mangoes, box g contains 1 apple, 8 oranges, and 4 mangoes, and box y contains 1 apple 9 oranges and 2 mangoes. If a box is chosen at random with probabilities $p(r) = 0.1$, $p(b) = 0.4$, $p(g) = 0.3$, $p(y) = 0.2$ and a piece of fruit is selected from the box (with equal probability of selecting any of the items in the box).

    (a) (6 points) What is the probability of selecting an orange?

(b) (9 points) If we observe that the selected fruit is in fact a mango, what is the probability that it came from the red box?

11. (10 points) We are given a set of data points $x_1, x_2, \ldots, x_n$ that are i.i.d. drawn from the density function:

$$f(x|\sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right), -\infty < x < \infty, \sigma > 0$$

Find the maximum likelihood estimate of σ.