Introduction to Machine Learning (CSCI-UA.473): Homework 1

Instructor: Lerrel Pinto

September 6, 2022

Submission Instructions

You must typeset the answers using LATEX and compile them into a single PDF file. Name the pdf file as ⟨Your-NetID⟩_hw1.pdf and the notebook containing the coding portion as ⟨Your-NetID⟩_hw1.ipynb. The PDF file should contain solutions to both the theory portion and the coding portion. Submit the files through the following Google Form - https://forms.gle/Vqj9ry6o3mqim6Hm6 The due date is **September 20, 2022, 11:59 PM**. You may discuss the questions with each other but each student must provide their own answer to each question.

Questions

Probability and Calculus

Question 1 (10 points)

Two players take turns trying to kick a ball into the net in soccer. Player 1 succeeds with probability 1/5 and Player 2 succeeds with the probability 1/4. Whoever succeeds first wins the game and the game is over. Assuming that Player 1 takes the first shot, what is the probability that Player 1 wins the game? Please derive your answer.

Ans:

P(player 1 win)= P (player 1 get 1st shot) +P (player 2 lose player 1 get 2nd shot) +P (player 2 lose player 1 get 3rd shot) + $\cdots$

$= \frac{1}{5} + \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{4} + \left(\frac{4}{5}\right)^2 \cdot \left(\frac{3}{4}\right)^2 \cdot \frac{1}{4} + \left(\frac{4}{5}\right)^3 \cdot \left(\frac{3}{4}\right)^3 \cdot \frac{1}{4} + \cdots$

This is an infinite GP series, we can use $S_\infty = \frac{a}{1-r}$ to calculate.

$$= \frac{\frac{1}{5}}{1 - \frac{12}{20}} = \frac{\frac{1}{5}}{\frac{8}{20}} = \frac{1}{2}$$

$\therefore$ The probability for player 1 to win is $\frac{1}{2}$.

Question 2 (10 points)

You know that 1% of the population have COVID. You also know that 90% of the people who have COVID get a positive test result and 10% of people who do not have COVID also test positive. What is the probability that you have COVID given that you tested positive?
Ans:
$P($ COVID $|$ tested positive $) = \frac{(COVID \cap testedpositive)}{P(testedpositive)}$

$$= \frac{1\% \times 90\%}{1\% \times 90\% + 10\% \times (1 - \%)}$$

$$= \frac{1}{12} = 0.083$$

Question 3 (10 points)

Let the function $f(x)$ be defined as:

$$f(x) = \begin{cases} 0 & for\ x < 0 \\ \frac{1}{(1+x)} & otherwise. \end{cases} \tag{1}$$

Is $f(x)$ a PDF? If yes, then prove that it is a PDF. If no, then prove that it is not a PDF.
Ans:
If $f(x)$ is a PDF, it needs to satisfy: 1) $f(x) \geqslant 0$. 2) $\int f(x) = 1$. 3) $P(A) = P(a \leqslant X \leq b) = \int_A f(x)dx$.
$\int f(x) = \int_a^b x \cdot f(x)dx = \int_0^\infty \frac{1}{1+x}dx = \ln|1+x|_0^\infty$. Since the result of $\ln|1+x|_0^\infty$ is diverge, which is $\neq 1$, the pdf is not defined.

Question 4 (10 points)

Assume that X and Y are two independent random variables and both have the same density function:

$$f(x) = \begin{cases} 2x & if \quad\quad 0 \leq x \leq 1 \\ 0 & otherwise \end{cases} \tag{2}$$

What is the value of $\mathbb{P}(X + Y \leq 1)$?
Ans:

$$f(x) = \begin{cases} 2x & if 0 \leqslant x \leqslant 1 \\ 0 & else \end{cases} ; f(y) = \{ 2\ y if 0 \leqslant y \leqslant 1 0 else$$

Since both x and Y are indepent, $f(x, y) = f(x) \cdot f(y)$.

$$\therefore f(x, y) = \{ 4\ xy 0 \leq x, y \leqslant 1 0 else P(X + Y \leqslant 1) = \int_0^1 \int_0^{1-y} 4xy \cdot dx \cdot dy$$

$$= \int_0^1 \frac{4y}{2} \cdot x^2 \Big|_0^{1-y} dy = \int_0^1 2y \cdot (1-y)^2 dy = \int_0^1 2y + 2y^3 - 4y^2 dy = 2 \cdot \frac{y^2}{2} + 2 \cdot \frac{y^4}{4} - 4 \cdot \frac{y^3}{3} \Big|_0^1 = 1 + \frac{1}{2} - \frac{4}{3} = \frac{1}{6}$$

Question 5 (10 points)

Let X be a random variable which belongs to a Uniform distribution between 0 and 1: $X \sim Unif(0,1)$. Let $Y = g(X) = e^X$. What is the value of $\mathbb{E}(Y)$?
Ans:

$$E(Y) = E(g(x)) = E\left(e^x\right) = \int_0^1 e^x \cdot 1dx = e^x\Big|_0^1 = e - 1 \approx 1.72 \therefore E(Y) = 1.72$$

Question 6 (10 points)

Suppose that the number of errors per computer program has a Poisson distribution with mean 5. We have 125 program submissions. Let $X_1, X_2, \ldots, X_{125}$ denote the number of errors in the programs. What is the value of $\mathbb{P}(\bar{X}_n < 5.5)$?
Ans:

$$E(\bar{X_n}) = E\left(\frac{1}{125}\sum_{i=1}^{125} X_i\right) = \frac{1}{125} \times 125 \times E\left(X_i\right) = 5$$

$$Var\left(\bar{X}_n\right) = Var\left(\frac{1}{125}\sum_{i=1}^{125} X_i\right) = \frac{1}{125 \times 125} \times 125 \times Var\left(X_i\right) = \frac{1}{125} \times 5 = \frac{1}{25}$$

By central limit theorem,

$$P\left(\bar{X}_n < 5.5\right) = P\left(\frac{\bar{X}_n - E\left(\bar{X}_n\right)}{\sqrt{Var\left(\bar{X}_n\right)}} < \frac{5.5-5}{\sqrt{\frac{1}{25}}}\right) = P\left(z < \frac{0.5}{\frac{1}{5}}\right) = P(z < 2.5) = 0.99$$

Question 7 (10 points)

Let $X_n = f(W_n, X_{n-1})$ for $n = 1, \ldots, P$, for some function $f()$. Let us define the value of variable E as

$$E = ||C - X_P||^2, \tag{3}$$

for some constant C. What is the value of the gradient $\frac{\partial E}{\partial X_0}$?
Since $E = \|C - X_p\|^2$,

$$\frac{\partial E}{\partial X_0} = \frac{\partial \left\|C - X_p\right\|^2}{\partial X_0} = \frac{\partial}{\partial X_p}\left\|C - X_p\right\|^2 \cdot \frac{\partial \left(C - X_p\right)}{\partial X_0} = -2\left(C - X_p\right) \cdot \frac{\partial X_p}{\partial X_0}$$

Since $X_n = f\left(W_n, X_{n-1}\right)$ for $n = 1, 2, \ldots, p$,

$$X_p = f\left(W_p, X_{p-1}\right).$$

So, $\frac{\partial X_p}{\partial X_0} = \frac{\partial f(W_p, X_{p-1})}{\partial X_{p-1}} \cdot \frac{\partial X_{p-1}}{\partial X_0}$. Similarly, $X_{p-1} = f\left(W_{p-1}, X_{p-2}\right)$.

$$\frac{\partial X_{p-1}}{\partial X_0} = \frac{\partial f\left(W_{p-1}, X_{p-2}\right)}{\partial X_{p-2}} \cdot \frac{\partial X_{p-2}}{\partial X_0} \therefore \frac{\partial X_p}{\partial X_0} = \frac{\partial f\left(W_p, X_{p-1}\right)}{\partial X_{p-1}} \cdot \frac{\partial X_{p-1}}{\partial X_0}$$

$$= \frac{\partial f\left(W_p, X_{p-1}\right)}{\partial X_{p-1}} \cdot \frac{\partial f\left(W_{p-1}, X_{p-2}\right)}{\partial X_{p-2}} \cdot \frac{\partial X_{p-2}}{\partial X_0}$$

Repeating this process, we will have

$$\frac{\partial X_p}{\partial X_0} = \frac{\partial f\left(W_p, X_{p-1}\right)}{\partial X_{p-1}} \cdot \frac{\partial f\left(W_{p-1}, X_{p-2}\right)}{\partial X_{p-2}} \cdot \cdots \cdot \frac{\partial f\left(W_2, X_1\right)}{\partial X_1} \cdot \frac{\partial f\left(W_1, X_0\right)}{\partial X_0}$$

$$\therefore \frac{\partial E}{\partial X_0} = -2\left(C - X_p\right) \cdot \frac{\partial X_p}{\partial X_0} = -2\left(C - X_p\right) \cdot \frac{\partial f\left(W_p, X_{p-1}\right)}{\partial X_{p-1}} \cdot \frac{\partial f\left(W_{p-1}, X_{p-2}\right)}{\partial X_{p-2}} \cdot \cdots \cdot \frac{\partial f\left(W_2, X_1\right)}{\partial X_1} \cdot \frac{\partial f\left(W_1, X_0\right)}{\partial X_0}$$

Linear Algebra

Question 8 (10 points)

Let A be the matrix $\begin{bmatrix} 2 & 6 & 7 \\ 3 & 1 & 2 \\ 5 & 3 & 4 \end{bmatrix}$ and let x be the column vector $\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$. Let A^T and x^T denote the transpose of A and x respectively. Compute Ax, A^T and $x^T A$.

Ans:

$$Ax = \begin{bmatrix} 2 & 6 & 7 \\ 3 & 1 & 2 \\ 5 & 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \cdot 2 + 6 \cdot 3 + 7 \cdot 4 \\ 2 \cdot 3 + 1 \cdot 3 + 2 \cdot 4 \\ 5 \cdot 2 + 3 \cdot 3 + 4 \cdot 4 \end{bmatrix} = \begin{bmatrix} 4 + 18 + 28 \\ 6 + 3 + 8 \\ 10 + 9 + 16 \end{bmatrix} = \begin{bmatrix} 50 \\ 17 \\ 35 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 2 & 6 & 7 \\ 3 & 1 & 2 \\ 5 & 3 & 4 \end{bmatrix}^T = \begin{bmatrix} 2 & 3 & 5 \\ 6 & 1 & 3 \\ 7 & 2 & 4 \end{bmatrix}$$

$$x^T A = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}^T \begin{bmatrix} 2 & 6 & 7 \\ 3 & 1 & 2 \\ 5 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 6 & 7 \\ 3 & 1 & 2 \\ 5 & 3 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \cdot 2 + 3 \cdot 3 + 4 \cdot 5 & 6 \cdot 2 + 1 \cdot 3 + 3 \cdot 4 & 7 \cdot 2 + 2 \cdot 3 + 4 \cdot 4 \end{bmatrix} = \begin{bmatrix} 33 & 27 & 36 \end{bmatrix}$$

Question 9 (10 points)

Find out if the following matrices are invertible. If yes, find the inverse of the matrix.

(a)

$$\begin{bmatrix} 6 & 2 & 3 \\ 3 & 1 & 1 \\ 10 & 3 & 4 \end{bmatrix} \tag{4}$$

Ans: Let the matrix be A. $det(A) = 6 \cdot (1.4 - 1 \cdot 3) - 2(3.4 - 1 \cdot 10) + 3(3 \cdot 3 - 1 \cdot 10) = 6 \cdot 1 - 2 \cdot 2 + 3(-1) = -1$ Since the determinant of the matrix is not equal to 0 , the matrix is invertible.

$$A^T = \begin{bmatrix} 6 & 3 & 10 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \end{bmatrix}$$

$$Adj(A) = \begin{bmatrix} 1 \times 4 - 1 \times 3 & -(2 \times 4 - 3 \times 3) & 2 \times 1 - 3 \times 1 \\ -(3 \times 4 - 1 \times 10) & 6 \times 4 - 3 \times 10 & -(1 \times 6 - 3 \times 3) \\ 3 \times 3 - 1 \times 10 & -(6 \times 3 - 2 \times 10) & 6 \times 1 - 2 \times 3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & -1 \\ -2 & -6 & 3 \\ -1 & 2 & 0 \end{bmatrix}$$

$$A^{-1} = \frac{1}{det(A)} \times Adj(A) = \frac{1}{-1} \times \begin{bmatrix} 1 & 1 & -1 \\ -2 & -6 & 3 \\ -1 & 2 & 0 \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 \\ 2 & 6 & -3 \\ 1 & -2 & 0 \end{bmatrix}$$

(b)

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 2 \\ 1 & 4 & 5 \end{bmatrix} \tag{5}$$

Ans:
Let the matrix be B.

$det(B) = 1 \cdot (2 \cdot 5 - 2 \cdot 4) - 2(0.5 - 1 \cdot 2) + 3(0.4 - 1 \cdot 2) = 1 \cdot 2 - 2 \cdot (-2) + 3 \cdot (-2) = 0$

Since the determinant of matrix is 0 , the matrix is not inverible.

Question 10 (10 points)

What is an Eigen Value of a matrix? What is an Eigen Vector of a matrix? Describe one method (any method) you would use to compute both of them.
Ans:
Let A be an $n \times n$ matrix and let $X \in \mathbb{C}^n$ be a nonzero vector for which

$$AX = \lambda X$$

for some scalar λ. Then λ is called an eigenvalue of the matrix A and X is called an eigenvector of A associated with λ, or a λ-eigenvector of A.

To compute eigenvalue, we should first calculate $A - \lambda I$. Then, we should find the determinant of $A - \lambda I$ and let it equal to zero. At this point, we will get a linear equation with one unknown variable λ. After we solve the equation what is/are λ , we could find out the eigenvalues. To find eigenvectors, we need to find out the eigenvector for each eigenvalue. First, we should use $A * \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ $=$ eigenvalue $* \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ to set up an equation. Then, extend each side of the equation we could get a set of equation containing x,y, and z. By solving this set of equation, we could get the answer for $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ which is our eigenvector. We should repeat this process to find all the eigenvectors for each eigenvalue.

Use the above described method to compute the Eigen Values of the matrix:

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & 0 \\ -2 & 2 & 1 \end{bmatrix} \tag{6}$$

Ans: $A - \lambda I =$

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & 0 \\ -2 & 2 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1-\lambda & 0 & -1 \\ 1 & -\lambda & 0 \\ -2 & 2 & 1-\lambda \end{bmatrix}$$

$$\begin{aligned} det(A - \lambda I) = (1-\lambda)[-\lambda \cdot (1-\lambda)] + (-1)(2-2\lambda) = 0 \\ (1-\lambda)\left(-\lambda + \lambda^2\right) - 2 + 2\lambda = 0 \\ \lambda_1 = 1, \lambda_2 = -1 \quad, \lambda_3 = 2 \end{aligned}$$

Thus, the eigenvalues of this matrix are $1, -1, 2$.