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.

Solution:
Player 1 needs Player 2 to not succeed until they do. So probability of Player 1 winning will be: $P_1 + ((P_1^{`}.P_2^{`}).P_1) + \left((P_1^{`}.P_2^{`})^2.P_1\right) + \left((P_1^{`}.P_2^{`})^3.P_1\right) + ...$ where P_1 indicates player 1 succeeding and $P_1^{`}$ indicates player 1 not succeeding in a

trial $(1 - P_1)$. So:

$P_1 = \frac{1}{5}, P_2 = \frac{1}{4}$

$P(\text{Player 1 Winning}) = P_1 + ((P_1^{`}.P_2^{`}).P_1) + \left((P_1^{`}.P_2^{`})^2.P_1\right) + \left((P_1^{`}.P_2^{`})^3.P_1\right) + ...$

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

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

$= \frac{1}{5}\frac{1}{1-\frac{3}{5}}$(From geometric series)

$\boxed{= \frac{1}{2}}$

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?

Solution:

$$P(C) = 0.01 \quad P(+ \mid C) = 0.9 \quad P(+ \mid \bar{C}) = 0.1$$

$$P(C \mid +) = \tfrac{P(+|C)P(C)}{P(+)}$$

$$= \tfrac{P(+|C)P(C)}{P(+|C)P(C)+P(+|\bar{C})P(\bar{C})}$$

$$= \tfrac{0.9\times0.01}{0.9\times0.01+0.1\times0.99} = \boxed{\frac{1}{12}}$$

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.

Solution: In order for a function to be a PDF the integral of the function should

$$\int_{-\infty}^{\infty} \frac{1}{1+x}dx = \int_{0}^{\infty} \frac{1}{1+x}dx = \ln(1+x)\Big]_0^{\infty} = \lim_{x\to\infty} \ln(1+x) = \infty$$

$\boxed{\text{So given function cannot be a PDF.}}$

Question 4 (10 points)

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

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

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

Solution:

$$f(x,y) = f(x)f(y) = \begin{cases} 4xy, 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1 \\ 0 \end{cases} \qquad (x,1)$$

$$P(x+y \leqslant 1) = \int_0^1 \int_0^{1-x} f(x,y)dydyx$$

$$= \int_0^1 \int_0^{1-x} 4xydydx.$$

First we'll solve the inner integral and then the outer integral as follows:

$$\int_0^{1-x} 4xy \cdot dy = \frac{4x}{2}y^2\bigg|_0^{1-x} = 2x(1-x)^2 - 2x \cdot 0$$

$$\int_0^1 2x(1-x)^2 dx = \int_0^1 \left(2x^3 - 4x^2 + 2x\right) dx$$

$$= \frac{x^4}{2} - \frac{4x^3}{3} + x^2\Big|_0^1$$

$$= \left(\frac{1}{2} - \frac{4}{3} + 1\right) - (0 - 0 + 0)$$

$$= \boxed{\frac{1}{6}}$$

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 $\mathbb{E}(Y)$?

Solution:

$$E[Y] = E\left[e^x\right]$$

$$= \int_{-\infty}^{\infty} e^x \cdot f_x(x) \cdot dx$$

$$= \int_0^1 e^x \cdot dx = \boxed{e-1}$$

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 $\mathbb{P}(\bar{X}_n < 5.5)$?

Solution: Solution comes from the Central Limit Theorem. (For reference please look into the 6th page of pdf: https://www.stat.cmu.edu/ larry/=stat325.01/chapter5.pdf).

In short, Central Limit Theorem states that if you have a population with a given mean μ and standard deviation σ and if we take large random samples from that population, mean of those samples will approximately normally distributed with mean μ and variance $\frac{\sigma^2}{n}$.

So we can say:

$$\begin{aligned}
&Z_n \sim N(0,1)\\
&\bar{X_n} \sim N(\mu, \frac{\sigma^2}{n})\\
&\bar{X_n} - \mu \sim N(0, \frac{\sigma^2}{n})\\
&\sqrt{n}(\bar{X_n} - \mu) \sim N(0, \sigma^2)\\
&\frac{\sqrt{n}(\bar{X_n} - \mu)}{\sigma} = Z_n \sim N(0,1)
\end{aligned}$$

In this setting if we're looking for probability where $\bar{X_n}$ is smaller than 5.5, then we can convert $\bar{X_n}$ to Z_n (a normal distribution $N(0,1)$) then find the possibilities in the new range.

$$P(\bar{X_n} \leq 5.5) = P\left(\frac{Z_n \sigma}{\sqrt{n}} + \mu \leq 5.5\right)$$

With Poisson distribution we know that $\mu = \sigma^2 = 5$ And $n = 125$.

$$\begin{aligned}
&= P\left(\frac{Z_n\sqrt{5}}{5\sqrt{5}} + 5 \leq 5.5\right)\\
&= P\left(\frac{Z_n}{5} \leq 0.5\right)\\
&= P(Z_n \leq 2.5)(\text{This is a known value from Normal Distribution})\\
&= \boxed{0.9938}
\end{aligned}$$

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 $\frac{\partial E}{\partial X_0}$?

Solution:

$$\frac{\partial X_n}{\partial X_{n-1}} = f'\left(W_n, X_{n-1}\right)$$

$$\frac{\partial E}{\partial X_0} = \underbrace{\frac{\partial E}{\partial X_p}}\underbrace{\frac{\partial X_p}{\partial X_{p-1}}\frac{\partial X_{p-1}}{\partial X_{p-2}} \cdots \frac{\partial X_1}{\partial X_0}}$$

$$= \boxed{-2\left(C - f\left(W_p, X_{p-1}\right)\right)\prod_{i=1}^{p}\frac{\partial f\left(W_i, X_{i-1}\right)}{\partial X_{i-1}}}$$

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 $x^T A$.

Solution:

$$\mathbf{Ax} = \begin{bmatrix} 2 & 6 & 7 \\ 3 & 1 & 2 \\ 5 & 3 & 4 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 50 \\ 17 \\ 35 \end{bmatrix}$$

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

$$\mathbf{x}^T\mathbf{A} = \begin{bmatrix} 2 & 3 & 4 \end{bmatrix}\begin{bmatrix} 2 & 6 & 7 \\ 3 & 1 & 2 \\ 5 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 33 & 27 & 36 \end{bmatrix}$$

Question 9 (10 points)

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

(a)

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

(b)

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

Solution:

(a) $\det(A) = (6 \times 1 \times 4) + (2 \times 1 \times 10) + (3 \times 3 \times 3) - (3 \times 1 \times 10) - (2 \times 3 \times 4) - (6 \times 1 \times 3) = 24 + 20 + 27 - 30 - 24 - 18 = -1 \neq 0$. So this matrix is invertible. A^{-1} is as follows:

$$\boxed{\mathbf{A}^{-1} = \begin{bmatrix} -1 & -1 & 1 \\ 2 & 6 & -3 \\ 1 & -2 & 0 \end{bmatrix}}$$

(b) $\det(B) = 0$. $\boxed{\text{Therefore, this matrix is NOT invertible}}$.

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

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

Solution:

Eigenvector is the vector of a linear transformation that only changes by a scalar factor when that transformation is applied to it. And eigenvalue is the scalar factor that eigenvector changed. If we refer to eigenvector as $\mathbf{v}$ and eigenvalue as λ for transformation $\mathbf{A}$ then we can write $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ for $\lambda \in \mathbb{R}$ and $\mathbf{v} \neq \mathbf{0}$. To find the eigenvalues and eigenvectors:

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0 \rightarrow \det(\mathbf{A} - \lambda\mathbf{I}) = 0$$

Then:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = -(\lambda - 1)(\lambda - 2)(\lambda + 1) = 0$$

There are 3 roots of this equation which corresponds to the eigenvalues of $\mathbf{A}$:

$$\boxed{\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = -1}$$

And in order to find the corresponding eigenvector for an eigenvalue we need to find $\mathbf{v}$ that satisfies $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0$. For λ_1 and $\mathbf{v}_1$:

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \begin{bmatrix} 1-1 & 0 & -1 \\ 1 & -1 & 0 \\ -2 & 2 & 1-1 \end{bmatrix} \begin{bmatrix} v_1^1 \\ v_1^2 \\ v_1^3 \end{bmatrix} = 0$$

And solve this multi variate equation for each eigenvector. Then we have the following eigenvectors:

$$\boxed{\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}}$$