HOMEWORK 1

>>Hemanth Sridhar Nakshatri<<
>>nakshatri@wisc.edu, 9085807346<<

Instructions: This is a background self-test on the type of math we will encounter in class. If you find many questions intimidating, we suggest you drop 760 and take it again in the future when you are more prepared.
Use this latex file as a template to develop your homework. Submit your homework on time as a single pdf file to Canvas. There is no need to submit the latex source or any code. Please check Piazza for updates about the homework.

1 Vectors and Matrices [6 pts]

Consider the matrix X and the vectors $\mathbf{y}$ and $\mathbf{z}$ below:

$$X = \begin{pmatrix} 3 & 2 \\ -7 & -5 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \qquad \mathbf{z} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

1. Compute $\mathbf{y}^T X \mathbf{z}$
   $y^T X z = \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ -7 & -5 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1 + 1 = 0$

2. Is X invertible? If so, give the inverse, and if no, explain why not.
   $X^{-1} = \frac{adj(X)}{det|X|}$

   $X^{-1} = \frac{\begin{bmatrix} -5 & -2 \\ 7 & 3 \end{bmatrix}}{-15+14} = \begin{bmatrix} 5 & 2 \\ -7 & -3 \end{bmatrix}$

   This results in $XX^{-1} = I$.
   Thus X in invertible.

2 Calculus [3 pts]

1. If $y = e^{-x} + \arctan(z)x^{6/z} - \ln \dfrac{x}{x+1}$, what is the partial derivative of y with respect to x?
   $\frac{dy}{dx} = -e^{-x} + \frac{6arctan(z)x^{(6/z-1)}}{z} - \frac{1}{x(x+1)}$

3 Probability and Statistics [10 pts]

Consider a sequence of data $S = (1, 1, 1, 0, 1)$ created by flipping a coin x five times, where 0 denotes that the coin turned up heads and 1 denotes that it turned up tails.

1. (2.5 pts) What is the probability of observing this data, assuming it was generated by flipping a biased coin with $p(x = 1) = 0.6$?

   $p(x = 1) = 0.6 \times 0.6 \times 0.6 \times 0.4 \times 0.6$
   $p(x = 1) = 0.05184$

2. (2.5 pts) Note that the probability of this data sample could be greater if the value of $p(x = 1)$ was not 0.6, but instead some other value. What is the value that maximizes the probability of S? Please justify your answer.
   Let $P(x = 1) = a$

Thus $P(S) = y = a^4(1-a)$
To maximize, $\frac{dy}{dx} = 0$

$\frac{dy}{da} = 4a^3 - 5a^4 = 0$
$\rightarrow 4a^3 = 5a^4$
$\rightarrow a = 4/5$

For maxima, $d^2y/dx^2 < 0$
$\frac{d^2y}{dx^2} = 12a^2 - 20a^3$ [Put a = 4/5]
$\rightarrow = -2.56 < 0$

Thus P(S) is maximum at $a = 4/5$ **or** 0.8

3. (5 pts) Consider the following joint probability table where both A and B are binary random variables:

A	B	$P(A,B)$
0	0	0.3
0	1	0.1
1	0	0.1
1	1	0.5

(a) What is $P(A = 0|B = 1)$?
$\rightarrow \frac{P(A=0)\cap P(B=1)}{P(B=1)} = \frac{0.1}{0.1+0.5}$
$\implies \frac{0.1}{0.6} = 1/6$

(b) What is $P(A = 1 \vee B = 1)$?
$\rightarrow P(A = 1) + P(B = 1) - P(A = 1 \cap B = 1)$
$\rightarrow 0.6 + 0.6 - 0.5 = \mathbf{0.7}$

4 Big-O Notation [6 pts]

For each pair (f, g) of functions below, list which of the following are true: $f(n) = O(g(n))$, $g(n) = O(f(n))$, both, or neither. Briefly justify your answers.

1. $f(n) = \ln(n)$, $g(n) = \log_2(n)$.
f(n) is *O(g(n))* if and only if there exists a positive 'c' and 'k' such that,
$|f(n)| \leq c.|g(n)|, for\ all\ n \geq k$

Choose k=2,
$ln(n) \leq c\ .\ log_2(n) for\ all\ n \geq 2$
$\frac{ln(n)}{log_2(n)} \leq c \implies log(2) \leq c$
$\implies f(n) \leq log(2)g(n)$
Thus, $f(n) = O(g(n))$

Alternatively we ccould use limits to check as well.
$L = \lim_{n\to\infty} \frac{g(n)}{f(n)} = \lim_{n\to\infty} \frac{\log_2(n)}{\ln(n)} = 1/\ln(2)$
Thus $g(n) = O(f(n))$.
Both statements are **True**

2. $f(n) = \log_2 \log_2(n)$, $g(n) = \log_2(n)$.
Choose k=3,

Similar to (1), we get,
$f(n) \leq \frac{log_2 log_2(n)}{log(n)}\ .\ g(n) for\ all\ n \geq 3$

Thus, $f(n) = O(g(n))$

3. $f(n) = n!$, $g(n) = 2^n$.
   $\rightarrow 2^n < n!$ for all $n \geq 4$
   for n = 4,
   $g(n) < c\,.\,f(n)$
   $\rightarrow c \geq 2/3$

   $\implies g(n) \leq \frac{2}{3}f(n) for\ all\ n \geq 4$

   Thus, $g(n) = O(f(n))$

5 Probability and Random Variables

5.1 Probability [12.5 pts]

State true or false. Here Ω denotes the sample space and A^c denotes the complement of the event A.

1. For any $A, B \subseteq \Omega$, $P(A|B)P(A) = P(B|A)P(B)$.
   FALSE

   $LHS \implies \frac{P(A \cap B)(P(A))}{P(B)}$
   $RHS \implies \frac{P(A \cap B)(P(B))}{P(A)}$
   Thus, it is FALSE

2. For any $A, B \subseteq \Omega$, $P(A \cup B) = P(A) + P(B) - P(B \cap A)$.
   TRUE, (Basic formula)

3. For any $A, B, C \subseteq \Omega$ such that $P(B \cup C) > 0$, $\frac{P(A \cup B \cup C)}{P(B \cup C)} \geq P(A|B \cup C)P(B)$.
   TRUE

   $\frac{P(A \cup B \cup C}{P(B \cup C)} \geq \frac{P(A \cap B \cup C)}{P(B \cup C)} P(B)$

   We know that *Max(P(B))* = 1. Thus maximum value of RHS is $\frac{P(A \cup B \cup C}{P(B \cup C)} \geq \frac{P(A \cap B \cup C)}{P(B \cup C}$.
   We also know that $P(X \cup Y) \geq P(X \cap Y)$
   $\implies P(A \cup B \cup C) \geq P(A \cap B \cup C)\ is$ **TRUE**

4. For any $A, B \subseteq \Omega$ such that $P(B) > 0, P(A^c) > 0$, $P(B|A^C) + P(B|A) = 1$.
   FALSE. (By total probability law, $P(B^c|A) + P(B|A) = 1$)

5. If A and B are independent events, then A^c and B^c are independent.
   TRUE

   If A and B are independent, then A^c and B^c are independent.

   $\implies P(A^c \cap B^c) = P((A \cup B)^c)$
   $= 1 - P(A \cup B)$
   $= 1 - P(A) - P(B) + P(A \cap B)$
   $= 1 - P(A) - P(B) + P(A)P(B)$
   $= (1 - P(A))(1 - P(B))$
   $= P(A^c)P(B^c)$

   TRUE

5.2 Discrete and Continuous Distributions [12.5 pts]

Match the distribution name to its probability density / mass function. Below, $|\boldsymbol{x}| = k$.

(a) Gamma **(j)**

(b) Multinomial **(i)**

(c) Laplace **(h)**

(d) Poisson **(l)**

(e) Dirichlet **(k)**

(f) $f(\boldsymbol{x}; \boldsymbol{\Sigma}, \boldsymbol{\mu}) = \frac{1}{\sqrt{(2\pi)^k \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$

(g) $f(x; n, \alpha) = \binom{n}{x}\alpha^x(1-\alpha)^{n-x}$ for $x \in \{0, \ldots, n\}$; 0 otherwise

(h) $f(x; b, \mu) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$

(i) $f(\boldsymbol{x}; n, \boldsymbol{\alpha}) = \frac{n!}{\Pi_{i=1}^k x_i!}\Pi_{i=1}^k \alpha_i^{x_i}$ for $x_i \in \{0, \ldots, n\}$ and $\sum_{i=1}^k x_i = n$; 0 otherwise

(j) $f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ for $x \in (0, +\infty)$; 0 otherwise

(k) $f(\boldsymbol{x}; \boldsymbol{\alpha}) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)}\prod_{i=1}^k x_i^{\alpha_i - 1}$ for $x_i \in (0, 1)$ and $\sum_{i=1}^k x_i = 1$; 0 otherwise

(l) $f(x; \lambda) = \lambda^x \frac{e^{-\lambda}}{x!}$ for all $x \in Z^+$; 0 otherwise

5.3 Mean and Variance [10 pts]

1. Consider a random variable which follows a Binomial distribution: $X \sim \text{Binomial}(n, p)$.

   (a) What is the mean of the random variable?
   $Mean = n \times p$

   (b) What is the variance of the random variable?
   $Var = np(1-p)$

2. Let X be a random variable and $\mathbb{E}[X] = 1, \text{Var}(X) = 1$. Compute the following values:

   (a) $\mathbb{E}[5X]$
   $\implies 5 \times E[X] = 5$

   (b) $\text{Var}(5X)$
   $\implies 5^2 Var(X) = 25$

   (c) $\text{Var}(X+5)$
   $\implies 1 \,.\, Var(X) = 1$

5.4 Mutual and Conditional Independence [12 pts]

1. (3 pts) If X and Y are independent random variables, show that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

   For independent varibales, w.k.t covariance(X,Y) = 0,
   The expectation of the product of two random variables,

   $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y] + Cov(X, Y)$
   $\implies \mathbb{E}[X] \cdot \mathbb{E}[Y] + 0$
   $\implies \mathbb{E}[X] \cdot \mathbb{E}[Y]$

2. (3 pts) If X and Y are independent random variables, show that $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.
   Hint: $\text{Var}(X+Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y)$

   $Var(X+Y) = Var(X) + 2Cov(X, Y) + Var(Y)$

   For independent variables, $Cov(X, Y) = 0$

   Thus, $Var(X+Y) = Var(X) + Var(Y)$

3. (6 pts) If we roll two dice that behave independently of each other, will the result of the first die tell us something about the result of the second die?

   No, the first die doesn't influence the second since they are independent.

   If, however, the first die's result is a 1, and someone tells you about a third event — that the sum of the two results is even — then given this information is the result of the second die independent of the first die?

With the given constraint, the second die is **NOT** independent of the first die since the first die result restricts the outcomes of the second die so that the sum is even.

5.5 Central Limit Theorem [3 pts]

Prove the following result.

1. Let $X_i \sim \mathcal{N}(0,1)$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, then the distribution of $\bar{X}$ satisfies

$$\sqrt{n}\bar{X} \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1)$$

Consider variable X_i with mean = 0 and varience, $\sigma^2 = 1$.
According to central limit theorem, addition of normal distribution, we get,
$n \cdot \bar{X} \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$

Normalizing the variance,
$\sqrt{n} \cdot \bar{X} \sim \mathcal{N}(\mu = 0, \sigma^2 = \sqrt{n})$

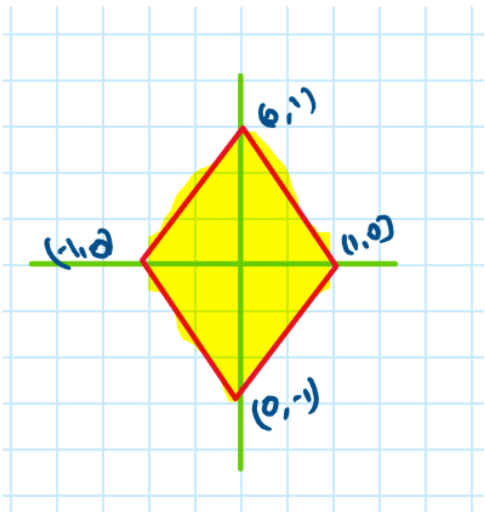
Also written as,
$\sqrt{n}(\bar{X}) \xrightarrow{n\to\infty} \mathcal{N}(0,1)$

6 Linear algebra

6.1 Norms [5 pts]

Draw the regions corresponding to vectors $\mathbf{x} \in \mathbb{R}^2$ with the following norms:

1. $||\mathbf{x}||_1 \leq 1$ (Recall that $||\mathbf{x}||_1 = \sum_i |x_i|$)
   Refer figure 1

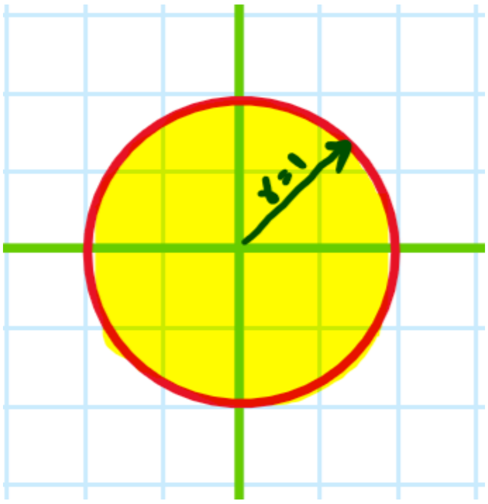


Plot for 6.1.1

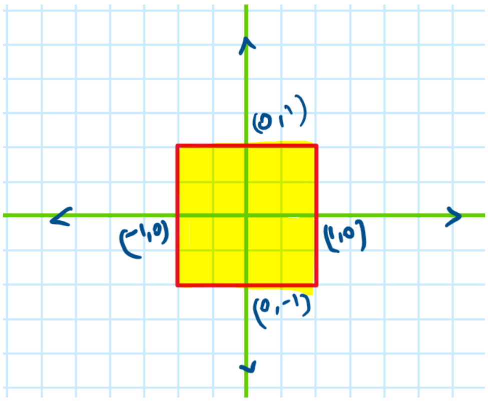
2. $||\mathbf{x}||_2 \leq 1$ (Recall that $||\mathbf{x}||_2 = \sqrt{\sum_i x_i^2}$) Refer figure 2

3. $||\mathbf{x}||_\infty \leq 1$ (Recall that $||\mathbf{x}||_\infty = \max_i |x_i|$) Refer figure 3.

For $M = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 3 \end{pmatrix}$, Calculate the following norms.



Plot for 6.1.2



Plot for 6.1.3

4. $||M||_2$ (L2 norm)
   M is a Diagonal matrix,
   L2 norm = 7

5. $||M||_F$ (Frobenius norm)

   $||M||_F = \sqrt{25 + 49 + 9} = 9.1104$

6.2 Geometry [10 pts]

Prove the following. Provide all steps.

1. The smallest Euclidean distance from the origin to some point $\mathbf{x}$ in the hyperplane $\mathbf{w}^T\mathbf{x} + b = 0$ is $\frac{|b|}{||\mathbf{w}||_2}$. You may assume $\mathbf{w} \neq 0$.
   $\implies w^T x + b = 0$
   $\implies w^T x = -b$

   Distance of a point on a hyperplane from the Origin is Projection of x drawn on the normal given by $|x \cdot a|$, where x is a point on the hyperplane and a denotes the unit normal of the plane.

   $\implies |x \cdot a| = \frac{|w^T x|}{||w||_2}$
   $\implies \frac{-b}{||w||_2} = \frac{|b|}{||w||_2}$

2. The Euclidean distance between two parallel hyperplane $\mathbf{w}^T\mathbf{x} + b_1 = 0$ and $\mathbf{w}^T\mathbf{x} + b_2 = 0$ is $\frac{|b_1 - b_2|}{||\mathbf{w}||_2}$ (Hint:

you can use the result from the last question to help you prove this one).

Similar to previous solution we have,
$w^T x = -b_1$
$w^T x = -b_2$

Distance between two hyperplanes is given by the projection of the difference vector $(x_1 - x_2)$ on the common unit normal.

Projection = $|(x_1 - x_2) \cdot a|$, where x_1 and x_2 are points on the hyperplane.

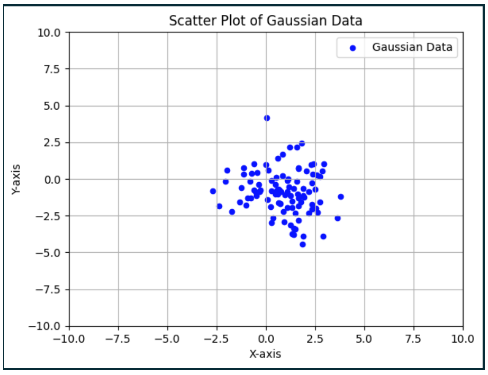
$\implies |(x_1 - x_2) \cdot a| = \frac{|(x_1 - x_2) w^T|}{||w||_2}$
$\implies \frac{|x_1 w^t - x_2 w^T|}{||w||_2} = \frac{|b_1 - b_2|}{||w||_2}$

7 Programming Skills [10 pts]

Sampling from a distribution. For each question, submit a scatter plot (you will have 2 plots in total). Make sure the axes for all plots have the same ranges.

1. Make a scatter plot by drawing 100 items from a two dimensional Gaussian $N((1, -1)^T, 2I)$, where I is an identity matrix in $\mathbb{R}^{2\times 2}$.

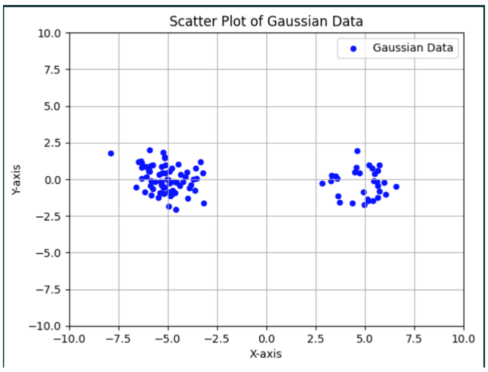
   See figure 4



Plot for 7.1

2. Make a scatter plot by drawing 100 items from a mixture distribution $0.3N\left((5,0)^T, \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}\right) + 0.7N\left((-5,0)^T, \begin{pmatrix} 1 & -0.25 \\ -0.25 & 1 \end{pmatrix}\right)$.

   See figure 5



Plot for 7.2