HOMEWORK 1

>>RISHIDEEP REDDY RALLABANDI<<
>>9084949099<<

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 $\mathbf{y}^T X \mathbf{z}$
   $\mathbf{y}^T X = \begin{pmatrix} -1 & -1 \end{pmatrix}$
   $\mathbf{y}^T X \mathbf{z} = \begin{pmatrix} 0 \end{pmatrix}$

2. Is X invertible? If so, give the inverse, and if no, explain why not.
   Yes it is invertible. It is a 2x2 matrix and the det(X) = -1. and the inverse is

$$X = \begin{pmatrix} 5 & 2 \\ -7 & -3 \end{pmatrix}$$

2 Calculus [3 pts]

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

$$y = -e^{-x} + \frac{6}{z}\arctan(z)x^{\frac{6}{z}-1} - \frac{1}{x(x+1)}$$

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 $p(x = 1) = 0.6$?

$$(0.6)^4 \cdot 0.4 = 0.05184$$

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's assume the maximum probability to be p. So, the probability of the data is $p^4(1-p)$. To maximize it, we differentiate with respect to p and set it equal to zero:

$$4p^3 - 5p^4 = 0$$

Solving for p, we find $p = \frac{4}{5}$ and $p = 0$. For $p = 0$, we get the minimum value, and for $p = \frac{4}{5}$, we get the maximum value.

Therefore, $p(x = 1) = 0.8$ to maximize the probability of S.

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 $P(A = 0|B = 1)$?

$$\frac{0.1}{0.1 + 0.5} = \frac{1}{6}$$

(b) What is $P(A = 1 \vee B = 1)$?

$$1 - P(A = 0, B = 0) = 1 - 0.3 = 0.7$$

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. $f(n) = \ln(n)$, $g(n) = \log_2(n)$.
I believe both statements are true.

Let $f(n) = \ln(n)$ and $g(n) = \log_2(n)$, which is approximately $1.44 \ln(2)$. We can observe that we can upper bound each one over the other by a constant greater than 2.

$$f(n) < g(n), \text{ so it is clear that } f(n) = O(g(n)) \forall n \geq e \text{ and } C \geq 1$$
$$g(n) < 2f(n) \Rightarrow g(n) = O(f(n)) \text{ for every } n > e \text{ and } C \geq 2$$

2. $f(n) = \log_2 \log_2(n)$, $g(n) = \log_2(n)$.
I believe $f(n) = O(g(n))$.

$$f(n) \leq c_0 \cdot \log_2(n) \cdot \log_2(n) \leq c_1 \cdot \log_2(n)$$

So, $f(n) = O(g(n))$.

3. $f(n) = n!$, $g(n) = 2^n$.
I believe $g(n) = O(f(n))$.

$$f(n) = n! = n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \ldots \cdot 1 \geq 2 \cdot 2 \cdot 2 \cdot 2 \ldots \forall n$$

So, $2^n \leq 3 \cdot n! \Rightarrow g(n) = O(f(n))$.

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
2. For any $A, B \subseteq \Omega$, $P(A \cup B) = P(A) + P(B) - P(B \cap A)$.
   True
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
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
5. If A and B are independent events, then A^c and B^c are independent.
   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?
   np
   (b) What is the variance of the random variable?
   np(1-p)
2. Let X be a random variable and $\mathbb{E}[X] = 1, \text{Var}(X) = 1$. Compute the following values:
   (a) $\mathbb{E}[5X]$
   5
   (b) $\text{Var}(5X)$
   25
   (c) $\text{Var}(X + 5)$
   1

5.4 Mutual and Conditional Independence [12 pts]

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

   The continuous random variables X and Y are considered independent when their joint probability density function $f_{XY}(x, y)$ can be expressed as the product of their individual probability density functions $f_X(x)$ and $f_Y(y)$, i.e., $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$.

   In such cases, the expected value (mean) of the product XY is given by:

$$
\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)xy\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x)f_Y(y)xy\,dx\,dy \\
&= \left(\int_{-\infty}^{\infty} f_X(x)x\,dx\right)\left(\int_{-\infty}^{\infty} f_Y(y)y\,dy\right) \\
&= E(X) \cdot E(Y)
\end{aligned}
$$

   This result shows that when X and Y are independent, the expected value of their product is equal to the product of their individual expected values, $E(X) \cdot E(Y)$.

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

   As we know $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$.

   If X and Y are independent random variables, their covariance is zero because $E[XY] = E[X]E[Y]$. Substituting this result into the hint, we have $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

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 it won't say, 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?

   No, it is not independent of the first die because they have a even sum condition to hold.

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

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

   Given:

$$
X_i \sim \mathcal{N}(0, 1), \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i
$$

   The expected value of $\bar{X}$ is:

$$
E[\bar{X}] = \frac{0}{n} = 0
$$

   The variance of $\bar{X}$ is:

$$
\mathrm{Var}[\bar{X}] = \frac{1}{n^2}(1 + 1 + \ldots + n \text{ times}) = \frac{1}{n}
$$

   Now consider $Z = \sqrt{n}\bar{X} = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}}$.

   The expected value of Z is:

$$
E[Z] = 0
$$

The variance of Z is:

$$\text{Var}(Z) = 1$$

Now, let's apply the moment generating function to Z:

$$M_Z(t) = M\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i(t)\right) = M\left(\sum_{i=1}^{n} X_i\left(\frac{t}{\sqrt{n}}\right)\right) = (M_X(t/\sqrt{n}))^n.$$

Since $X_i \sim N(0,1)$, we know $M_X(t) = e^{t^2/2}$. Thus,

$$(M_X(t/\sqrt{n}))^n = \left(e^{(t/\sqrt{n})^2/2}\right)^n = e^{t^2/2}.$$

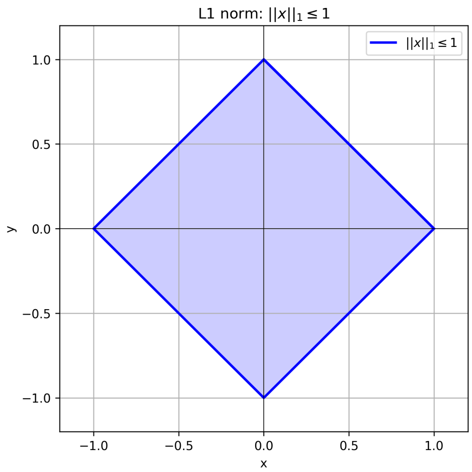
The moment generating function of Z_n is equal to the moment generating of standard guassian $\mathcal{N}(0,1)$. Therefore, as $n \to \infty$, $Z_n = \sqrt{n}\bar{X}$ converges in distribution to $\mathcal{N}(0,1)$.

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 $||\mathbf{x}||_1 = \sum_i |x_i|$) L1 Plot [1]

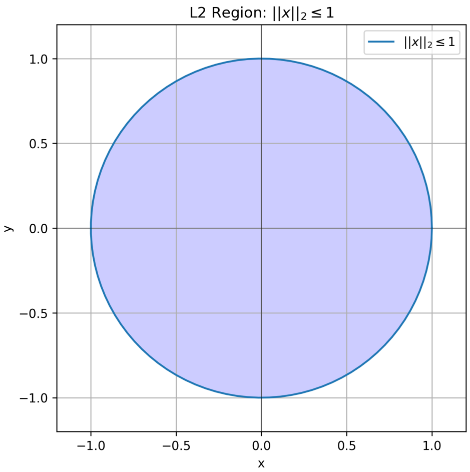


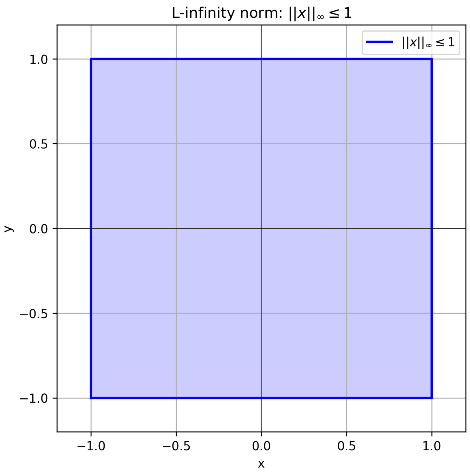
2. $||\mathbf{x}||_2 \leq 1$ (Recall that $||\mathbf{x}||_2 = \sqrt{\sum_i x_i^2}$) 2 L2 Plot [2]
3. $||\mathbf{x}||_\infty \leq 1$ (Recall that $||\mathbf{x}||_\infty = \max_i |x_i|$)
   Infinity Plot [3]

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

4. $||M||_2$ (L2 norm)
   7

5. $||M||_F$ (Frobenius norm)
   $\sqrt{83} \approx 9.11$





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 $\mathbf{w} \neq 0$.

   Let x be some point on the hyperplane $w^T x + b = 0$, where w is the normal vector. To compute the Euclidean distance from the origin to this hyperplane, we need to project x onto the normal vector of the hyperplane, denoted by w. This projection allows us to calculate the distance as follows:

$$\text{Distance} = \frac{|w^T x|}{\|w\|_2} = \frac{|-b|}{\|w\|_2} = \frac{|b|}{\|w\|_2}$$

   This calculation involves the dot product of w and x divided by the L2 norm of w, effectively measuring the distance from the origin to the hyperplane.

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).

   Let x_1 be a point on the first plane, and x_2 be a point on the second plane. When calculating the Euclidean distance, we must project the vector $x_1 - x_2$ onto the normal vector, following a similar procedure as previously explained.

   The distance can be computed as:

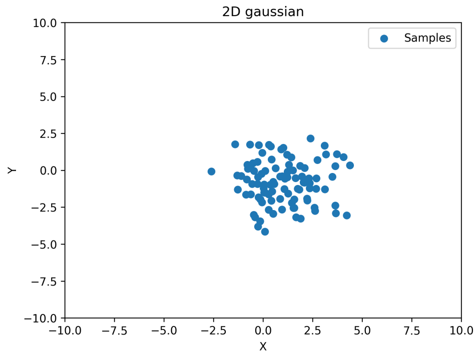
$$\text{Distance} = \frac{|w^T(x_1 - x_2)|}{\|w\|_2} = \frac{|b_1 - b_2|}{\|w\|_2}$$

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 $\mathbb{R}^{2\times 2}$.

   2D gaussian plot 4



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

   Multi gaussian plot 5