HOMEWORK 1

>>JIAHUI ZHANG<<
>>908 449 6323<<

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. Computer $\mathbf{y}^T X \mathbf{z}$
   Solution:
$$\begin{aligned} \mathbf{y}^T X \mathbf{z} =& \begin{pmatrix} -1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ =& 0 \end{aligned} \tag{1}$$

2. Is X invertible? If so, give the inverse, and if no, explain why not.
   Solution: Yes. The inverse is $\begin{pmatrix} 5 & 2 \\ -7 & -3 \end{pmatrix}$.

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?
   Solution:
$$\begin{aligned} d\frac{y}{dx} =& -e^{-x} + \arctan(z)\frac{6}{z}x^{6/z-1} - \frac{x+1}{x}\frac{1}{(x+1)^2} \\ =& -e^{-x} + \arctan(z)\frac{6}{z}x^{6/z-1} - \frac{1}{x(x+1)} \end{aligned} \tag{2}$$

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 $p(x = 1) = 0.6$?
   Solution:
$$\mathbb{P}(S) = 0.6^3 \times 0.4 \times 0.6 = 0.05184 \tag{3}$$

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.
Denote the probability of $p(x = 1)$ as p. Then the probability of seeing sequence S is $f(p) = p^4(1 - p)$. Then let

$$d\frac{f(p)}{p} = -p^4 + 4p^3(1 - p) = 0 \Rightarrow -p + 4(1 - p) = 0 \tag{4}$$

We obtain $p^* = 0.8$. And we have $f(p^*) = 0.08192$.

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

$$P(A = 0 \mid B = 1) = \frac{0.1}{0.1 + 0.5} = \frac{1}{6} \tag{5}$$

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

$$P(A = 1 \vee B = 1) = 0.1 + 0.1 + 0.5 = 0.7 \tag{6}$$

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. $f(n) = \ln(n)$, $g(n) = \log_2(n)$.
Solution:

$$\begin{aligned} \frac{f(n)}{g(n)} &= \frac{\log_e(n)}{\log_2(n)} \\ &= \frac{\log_2(n)/\log_2(e)}{\log_2(n)} \\ &= \frac{1}{\log_2(e)} \end{aligned} \tag{7}$$

Therefore, there exists constant $M = \frac{1}{\log_2(e)}$ such that $f(n) \leq Mg(n)$ for all $n > n_0$. And vice versa ($M = \log_2(e)$). Therefore, both $f(n) = O(g(n))$ and $g(n) = O(f(n))$ are true.

2. $f(n) = \log_2 \log_2(n)$, $g(n) = \log_2(n)$.
Solution:

$$\begin{aligned} f(n) &= \log_2(g(n)) \\ n \to \infty, g(n) \to \infty, &\frac{f(g(n))}{g(n)} \to 0 \end{aligned} \tag{8}$$

Therefore, $f(n) = O(g(n))$, but the reverse is not true.

3. $f(n) = n!$, $g(n) = 2^n$.
Solution:

$$\begin{aligned} \lim_{n\to\infty} \frac{f(n)}{g(n)} &= \lim_{n\to\infty} \frac{n!}{2^n} \\ &= \lim_{n\to\infty} \frac{(n/e)^n \sqrt{2\pi n}(1+\frac{1}{12n})}{2^n} \\ &= \infty \end{aligned} \tag{9}$$

Therefore, $g(n) = O(f(n))$ but the reverse is not true.

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 $A, B \subseteq \Omega$, $P(A|B)P(A) = P(B|A)P(B)$.
Solution:

$$\begin{aligned} P(A|B)P(A) &= \frac{P(A\cap B)P(A)}{P(B)} \\ P(B|A)P(B) &= \frac{P(A\cap B)P(B)}{P(A)} \end{aligned} \tag{10}$$

Therefore, if the statement holds true, we must have $P(A)^2 = P(B)^2$. Obviously this is not necessarily true for any $A, B \subseteq \Omega$. So the statement is false.

2. For any $A, B \subseteq \Omega$, $P(A \cup B) = P(A) + P(B) - P(B \cap A)$.
Solution: This is true.

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

$$\begin{aligned} P(A|B\cup C)P(B) &= \frac{P(A\cap(B\cup C))}{P(B\cup C)}P(B) \\ &\leq \frac{P(B\cup C)P(B)}{P(B\cup C)} \\ &\leq \frac{P(B\cup C)}{P(B\cup C)} \\ &\leq \frac{P(A\cup B\cup C)}{P(B\cup C)} \end{aligned} \tag{11}$$

Therefore the statement is true.

4. For any $A, B \subseteq \Omega$ such that $P(B) > 0, P(A^c) > 0$, $P(B|A^C) + P(B|A) = 1$.
Solution:

$$P(B|A^c) + P(B|A) = \frac{P(B\cap A^c)}{P(A^c)} + \frac{P(B\cap A)}{P(A)} \tag{12}$$

Without loss of generality, assume that $P(A) \geq P(A^c)$. Therefore

$$\frac{P(B\cap A^c)}{P(A^c)} + \frac{P(B\cap A)}{P(A)} \leq \frac{P(B\cap A^c) + P(B\cap A)}{P(A^c)} = \frac{P(B)}{P(A^c)} \tag{13}$$

And one can easily give examples on $P(B) < P(A^c)$. Therefore, the original expression can be less than 1. The statement is not true.

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

Solution: Independence means

$$\begin{aligned} P(A \cap B) =& P(A)P(B) = (1 - P(A^c))(1 - P(B^c)) \\ \therefore P(A^c \cap B^c) =& 1 - P(A \cup B) \\ =& 1 - P(A) - P(B) + P(A \cap B) \\ =& 1 - P(A) - P(B) + 1 - P(B^c) - P(A^c) + P(A^c)P(B^c) \\ =& 1 - 1 - 1 + 1 + P(A^c)P(B^c) \\ =& P(A^c)P(B^c) \end{aligned} \tag{14}$$

Therefore, the statement is 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 (g)
(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) $\mathbb{E}[5X]$
   $\mathbb{E}(5X) = 5\mathbb{E}(X) = 5$
   (b) $\text{Var}(5X)$
   $\text{Var}(5X) = 5^2\text{Var}(X) = 25$
   (c) $\text{Var}(X + 5)$
   $\text{Var}(X + 5) = \text{Var}(X) = 1$

5.4 Mutual and Conditional Independence [12 pts]

1. (3 pts) If X and Y are independent random variables, show that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
   Solution: When X and Y are independent, we have $f(x, y) = f(x)f(y)$. Then

$$\mathbb{E}(XY) = \int_x \int_y xyf(x, y)dydx = \int_x \int_y xyf(x)f(y)dydx = \mathbb{E}(X)\mathbb{E}(Y) \tag{15}$$

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

   Solution: Given the hint, we only need to prove that $\mathrm{Cov}(X,Y) = 0$. We have

   $$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0 \tag{16}$$

   Obtained from the last question. Then the statement is proved.

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. 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. They are no longer independent.

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

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

   Solution: Define $S_n = \sum_{i=1}^n X_i$, then we have that

   $$\begin{aligned}\sqrt{n}\bar{X} =& \frac{1}{\sqrt{n}} S_n \\ \because M_{S_n}(t) =& (M_x(t))^n \\ \therefore M_{\sqrt{n}\bar{X}}(t) =& \left(M_x(\frac{t}{\sqrt{n}})\right)^n\end{aligned} \tag{17}$$

   By Taylor's expansion, we have

   $$\begin{aligned}M_x(s) =& M_x(0) + sM_x'(0) + \frac{1}{2}s^2 M_x''(0) + o(s^2) \\ =& 1 + s\mathbb{E}(X_i) + \frac{1}{2}s^2\left(\sigma^2 + (M_x'(0))^2\right) + o(s^2) \\ =& 1 + \frac{s^2}{2} + o(s^2) \\ \therefore M_{\sqrt{n}\bar{X}}(t) =& \left(1 + \frac{t^2}{2n} + o(n)\right)^n \\ \therefore \lim_{n\to\infty} M_{\sqrt{n}\bar{X}}(t) =& \lim_{n\to\infty}\left(1 + \frac{t^2/2 + no(n)}{n}\right)^n = e^{\frac{t^2}{2}} \\ \therefore \sqrt{n}\bar{X} \xrightarrow{\mathrm{d}}& \mathcal{N}(0,1)\end{aligned} \tag{18}$$

6 Linear algebra

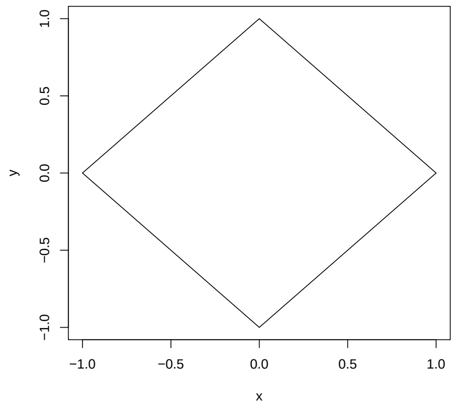
6.1 Norms [5 pts]

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

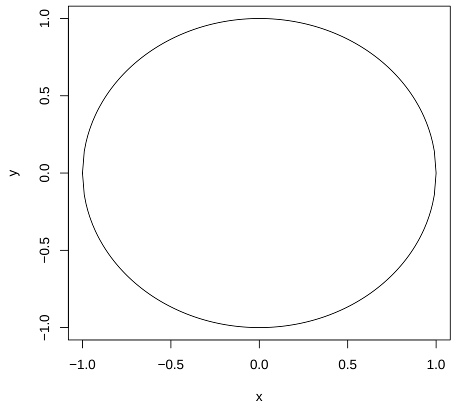
1. $||\mathbf{x}||_1 \le 1$ (Recall that $||\mathbf{x}||_1 = \sum_i |x_i|$)

   Solution: see Figure 1.

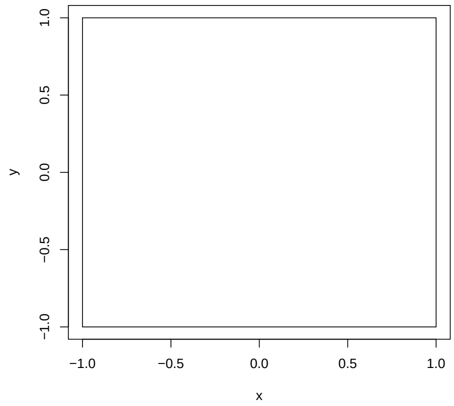
2. $||\mathbf{x}||_2 \le 1$ (Recall that $||\mathbf{x}||_2 = \sqrt{\sum_i x_i^2}$) Solution: see Figure 2.



6.1-1



6.1-2



6.1-3

3. $||\mathbf{x}||_\infty \leq 1$ (Recall that $||\mathbf{x}||_\infty = \max_i |x_i|$) Solution: see Figure 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)
   Solution: this is the largest singular value, which is 7.

5. $||M||_F$ (Frobenius norm)
   Solution:
$$||M||_F = \sqrt{\sum_i \sum_j |a_{ij}|^2} = \sqrt{5^2 + 7^2 + 3^2} = 9.11 \tag{19}$$

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 $\mathbf{w} \neq 0$.
   Solution: We have
$$\begin{aligned} \mathbf{w}^T\mathbf{x} &= \cos(\mathbf{w},\mathbf{x})||\mathbf{w}||_2||\mathbf{x}||_2 = -b \\ \because ||\mathbf{w}||_2 &> 0, ||\mathbf{x}||_2 \geq 0, \cos(\mathbf{w},\mathbf{x}) \in [-1,1] \\ \therefore ||\mathbf{x}||_2 &= \frac{-b}{||\mathbf{w}||_2 \cos(\mathbf{w},\mathbf{x})} \\ \therefore ||\mathbf{x}||_2 &\geq \frac{|b|}{||\mathbf{w}||_2} \end{aligned} \tag{20}$$

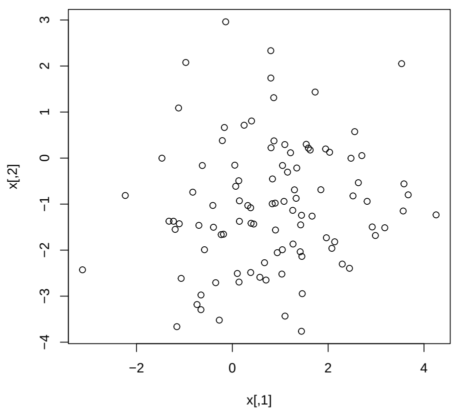
2. The Euclidean distance between two parallel hyperplane $\mathbf{w}^T\mathbf{x} + b_1 = 0$ and $\mathbf{w}^T\mathbf{x} + b_2 = 0$ is $\frac{|b_1 - b_2|}{||\mathbf{w}||_2}$ (Hint: you can use the result from the last question to help you prove this one).
   Solution: suppose $\mathbf{x}_1$ is on the first hyperplane and $\mathbf{x}_2$ is on the other. Then
$$\begin{aligned} \mathbf{w}^T\mathbf{x}_1 &= -b_1 \\ \mathbf{w}^T\mathbf{x}_2 &= -b_2 \\ \therefore \mathbf{w}^T(\mathbf{x}_1 - \mathbf{x}_2) &= b_2 - b_1 \end{aligned} \tag{21}$$
   Plugging in the derivation of the previous question, we have
$$||\mathbf{x}_1 - \mathbf{x}_2|| \geq \frac{|b_2 - b_1|}{||\mathbf{w}||_2} \tag{22}$$

7 Programming Skills [10 pts]

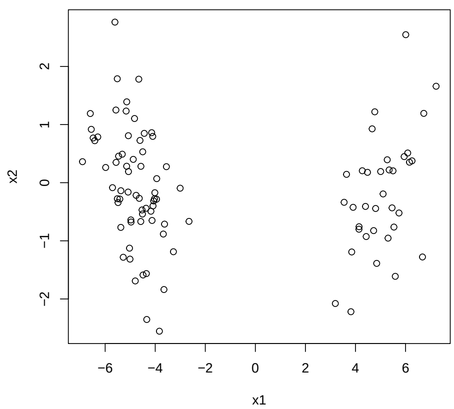
Sampling from a distribution. For each question, submit a scatter plot (you will have 2 plots in total). Make sure the axes for all plots have the same ranges.

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 $\mathbb{R}^{2\times 2}$.
   Solution: see Figure 4. I used package 'MASS' in R for this and the next question.

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



7.1



7.2