Homework 1

RYAN YEE
9074025223

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 $\mathbf{y}^T X \mathbf{z}$

$$
\begin{aligned}
\mathbf{y}^T X \mathbf{z} &= \begin{pmatrix} 2 & 1 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ -7 & -5 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
&= \begin{pmatrix} 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} \\
&= 0
\end{aligned}
$$

2. Is X invertible? If so, give the inverse, and if no, explain why not.

$$
\begin{aligned}
X^{-1} &= \frac{1}{-1} \begin{pmatrix} -5 & -2 \\ 7 & 3 \end{pmatrix} \\
&= \begin{pmatrix} 5 & 2 \\ -7 & -3 \end{pmatrix}
\end{aligned}
$$

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?

$$
\begin{aligned}
\frac{\partial y}{\partial x} &= -e^{-x} + \frac{6}{z}\arctan(z)x^{(6/z)-1} - \frac{x+1}{x}\left(\frac{(x+1)-x}{(x+1)^2}\right) \\
&= -e^{-x} + \frac{6}{z}\arctan(z)x^{(6/z)-1} - \frac{1}{x(x+1)}
\end{aligned}
$$

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 $p(x = 1) = 0.6$?

   $p(S) = 0.6^4 * (1 - 0.6) = 0.05184$

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.

   $p(x = 1) = 0.8$ maximizes the probability of S because it is the observed probability of success based on the observed data.

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

   $$P(A = 0|B = 1) = \frac{P(A = 0 \cap B = 1)}{P(B = 1)}$$
   $$= \frac{0.5}{0.1 + 0.5}$$
   $$= \frac{5}{6}$$
   $$\approx .8333333$$

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

   $$P(A = 1 \vee B = 1) = 1 - P(A = 0 \cap B = 0)$$
   $$= 1 - 0.3$$
   $$= 0.7$$

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. $f(n) = \ln(n)$, $g(n) = \log_2(n)$.
   Both $f(n) = O(g(n))$ and $g(n) = O(f(n))$ are true.
   $f'(n) = n^{-1}$ and $g'(n) = n^{-1} \ln(2)$. Therefore, by L'Hopital's rule, the limits of both $\frac{f(n)}{g(n)}$ and $\frac{g(n)}{f(n)}$ converge to a constant.

2. $f(n) = \log_2 \log_2(n)$, $g(n) = \log_2(n)$.
   Only $f(n) = O(g(n))$.
   We have, $f'(n) = n^{-1} \frac{1}{\log_2(n)} \ln(2)^{-2}$ and $g'(n) = n^{-1} \ln(2)$. So $\frac{f(n)}{g(n)}$ converges to zero but $\frac{g(n)}{f(n)} = \log_2(n)$.

3. $f(n) = n!$, $g(n) = 2^n$.
   Only $f(n) = O(g(n))$.
   As n approaches infinity, $n!$ grows much faster than 2^n. This is clear because $g(n + 1) = 2g(n)$ while $f(n + 1) = nf(n)$.

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, $|\boldsymbol{x}| = k$.

(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 $x \in \{0, \dots, n\}$; 0 otherwise

(a) Gamma (j)

(b) Multinomial (f)

(c) Laplace (h)

(d) Poisson (l)

(e) Dirichlet (k)

(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 $x_i \in \{0, \dots, n\}$ 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?
       $\mathbb{E}[X] = np$

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

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 $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} XY f_{XY} \partial X \partial Y = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} XY f_X f_Y \partial X \partial Y$$
$$= \int_{-\infty}^{\infty} X f_X \partial X \int_{-\infty}^{\infty} Y f_Y \partial Y$$
$$= \mathbb{E}[X]\mathbb{E}[Y]$$

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

   Since X and Y are independent, $\mathrm{Cov}(X, Y) = 0$.

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + 2\mathrm{Cov}(X, Y) + \mathrm{Var}(Y)$$
$$= \mathrm{Var}(X) + \mathrm{Var}(Y)$$

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?

   The result of the first die will not tell us anything about the result of the second die.

   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?

   Given this information, the result of the second die is not independent of the first. In this case, we know the second die must be odd since the first die was 1 and the sum of the die 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

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

   Show the mgf of $\sqrt{n}\bar{X}$ converges to the mgf of $\mathcal{N}(0, 1)$ which is $e^{t^2/2}$.
   Using properties of mgfs, we have $M_{\sqrt{n}\bar{X}}(t) = M_{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i}(t) = M_{\sum_{i=1}^{n} X_i}(\frac{t}{\sqrt{n}}) = (M_X(\frac{t}{\sqrt{n}}))^n$.
   Since $X_i \sim \mathcal{N}(0, 1)$ we know $M_X(t) = e^{t^2/2}$. Thus, $(M_X(\frac{t}{\sqrt{n}}))^n = (e^{(\frac{t}{\sqrt{n}})^2/2})^n = (e^{\frac{t^2}{n}/2})^n = e^{t^2/2}$.
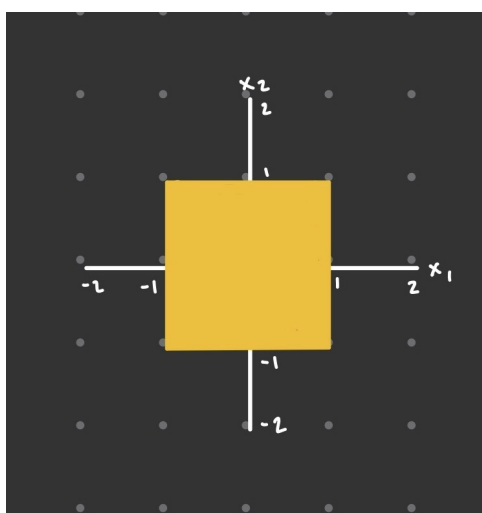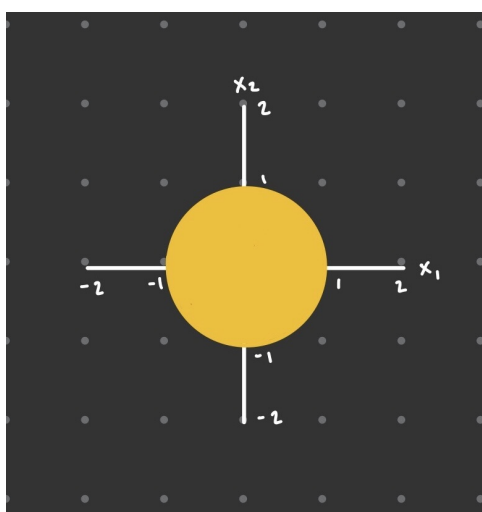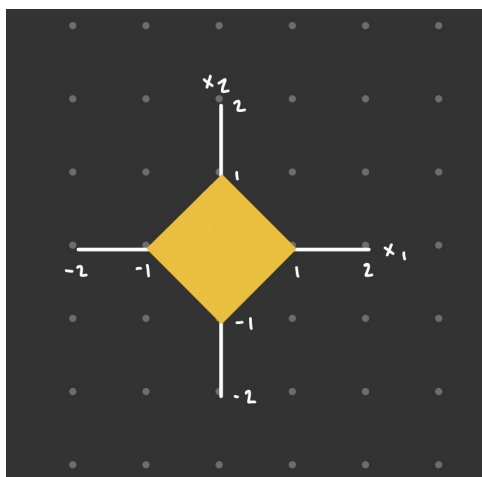
6   Linear algebra

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 $||\mathbf{x}||_1 = \sum_i |x_i|$)

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

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

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

4. $||M||_2$ (L2 norm)
   $||M||_2 = \sqrt{\lambda_{max}(M^T M)} = 7$

5. $||M||_F$ (Frobenius norm)
   $||M||_F = \sqrt{\mathrm{trace}(M^T M)} = \sqrt{83} \approx 9.11$

6.2   Geometry [10 pts]

Prove the following. Provide all steps.

5

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 any point in the hyperplane. Because w is orthogonal to the hyperplane, projecting z onto w will give a vector orthogonal to the hyperplane pointing from the origin to a point on the hyperplane, the length of this vector will be the smallest distance to the hyperplane. Let v be such vector. Then, $v = \frac{x \cdot w}{w \cdot w} w = \frac{-b}{w \cdot w} w$. Let d be the smallest distance from the origin to a point on the hyperplane. Then, $d = ||v||_2 = \frac{|b| ||w||_2}{||w||_2^2} = \frac{|b|}{||w||_2}$.

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).

   Since both hyperplanes are orthogonal to w, the distance between them is equivalent to the distances between the points closest to the origin. From above, we know the distance from the hyperplanes to the origin are $\frac{|b_i|}{||w||_2}$ where $i = 1, 2$. Finally, we either subtract the distances if the hyperplanes are in the same direction or add them if they are in opposite directions from the origin, which is equivalent to $\frac{|b_1 - b_2|}{||w||_2}$.

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

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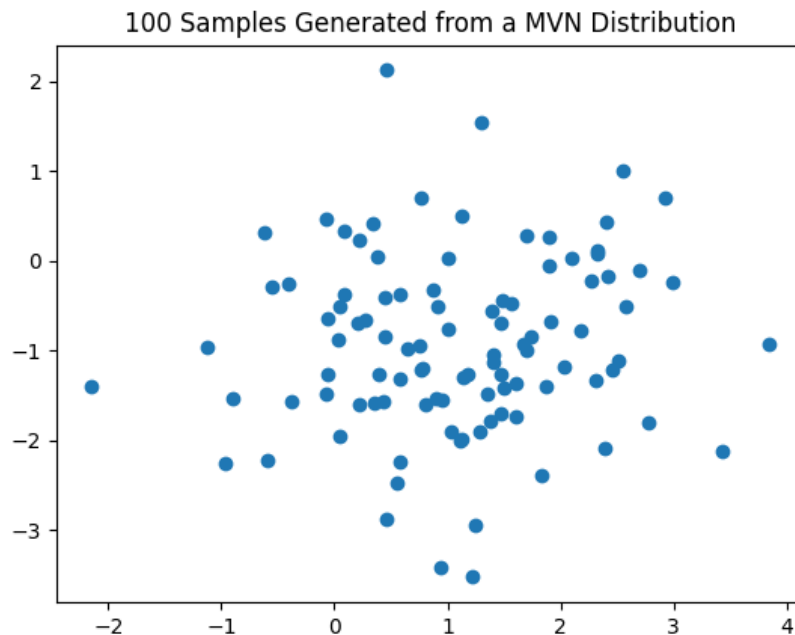
2. Make a scatter plot by drawing 100 items from a mixture distribution $0.3N\left((5, 0)^T, \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}\right) + 0.7N\left((-5, 0)^T, \begin{pmatrix} 1 & -0.25 \\ -0.25 & 1 \end{pmatrix}\right)$.

100 Samples Generated from a Gaussian Mixture