HOMEWORK 1

Name: Lucas Poon
ID: llpoon

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

2. Is X invertible? If so, give the inverse, and if no, explain why not.
   $\det(X) = 3*(-5) - 2*(-7) = -1$
   $X^{-1} = \frac{1}{-1} \begin{pmatrix} -5 & -2 \\ 7 & 3 \end{pmatrix} = \begin{pmatrix} 5 & 2 \\ -7 & -3 \end{pmatrix}$

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?

   $\frac{dy}{dx} = -e^{-x} + \frac{6 \arctan(z) x^{\frac{6}{z}-1}}{z} - \frac{1}{x^2+x}$

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 $p(x = 1) = 0.6$?
   $(0.6)^4 * 0.4 = 0.05184$

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 $P(S)$ denote the probability of the sequence S and p denote $p(x = 1)$, $0 \leq p \leq 1$,
   Then, $P(s) = p^4(1 - p)$, we want to find some p that maximize P(S).
   Differentiate to get the extreme values, $\frac{dP(S)}{dp} = p^3(4 - 5p)$,

$\frac{dP(S)}{dp} = 0$, when $p = 0$ or $p = \frac{4}{5}$.
We ignore $p = 0$, as we cannot get $x = 0$ if $p = 1$.
By taking the second derivative we can check if $p = \frac{4}{5}$ is a maximum, $\frac{d^2P(S)}{dp^2} = 12p^2 - 20p^3$.
$\because 12 * (\frac{4}{5})^2 - 20 * (\frac{4}{5})^3 = -64/25 < 0$,
$\therefore P(S)$ is greatest when $p = \frac{4}{5}$.

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 $P(A = 0|B = 1)$?
   $\frac{0.1}{0.6} = 0.1\dot{6}$
   (b) What is $P(A = 1 \vee B = 1)$?
   $0.6 + 0.6 - 0.5 = 0.7$

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 are true.
   $\log_2(n) = \frac{\ln(n)}{\ln(2)}, \ln(n) \leq \ln(2) * \frac{\ln(n)}{\ln(2)}, \forall n \geq 1$
   $\ln(n) = \frac{\log_2 n}{\log_2 e}, \log_2 n \leq \log_2 e * \frac{\log_2 n}{\log_2 e}, \forall n \geq 1$
2. $f(n) = \log_2 \log_2(n)$, $g(n) = \log_2(n)$.
   Only true for $f(n) = O(g(n)), \log_2(log_2 n) \leq \log_2 n, \forall n \geq 1$
   False for $g(n) = O(f(n)), \log_2 n \leq \log_2(log_2 n), \forall n \geq 1$
3. $f(n) = n!$, $g(n) = 2^n$.
   Only true for $g(n) = O(f(n)), 2^n \leq 3 * n!, \forall n \geq 1$
   False for $f(n) = O(g(n))$. $\because \frac{2^n}{n!} \underset{n\to\infty}{\to} 0$, we know that $n! > 2^n$

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
   Let D denote $B \cup C$, then we need to show that $\dfrac{P(A \cup D)}{P(D)} \geq P(A|D)P(B)$
   $\rightarrow \dfrac{P(A \cup D)}{P(D)} \geq \dfrac{P(A \cap D)}{P(D)} P(B)$
   $\rightarrow P(A \cup D) \geq P(A \cap D)P(B)$
   $\because P(A \cup D) \geq P(A \cap D)$ and $P(B) \leq 1$ $\therefore$ this statement holds.

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 $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
   For discrete variables X and Y,
   $\mathbb{E}[XY] = \sum_x \sum_y xyP(X = x, Y = y) = \sum_x \sum_y xyP(X = x)P(Y = y)$
   $= (\sum_x xP(X = x))(\sum_y yP(Y = y)) = \mathbb{E}[X]\mathbb{E}[Y]$

For continuous random variables X and Y,
$\mathbb{E}[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyP(X = x, Y = y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyP(X = x)P(Y = y)dxdy$
$= (\int_{-\infty}^{\infty} xP(X = x)dx)(\int_{-\infty}^{\infty} yP(Y = y)dy) = \mathbb{E}[X]\mathbb{E}[Y]$

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

   $\because$ X and Y are independent random variables, the $\text{Cov}(X, Y)$ is always zero.
   $\therefore$ $\text{Var}(X + Y) = \text{Var}(X) + 2(0) + \text{Var}(Y) = \text{Var}(X) + \text{Var}(Y)$

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, because 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 dependent on the first die.

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

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

$\sqrt{n}\bar{X} = \dfrac{\sqrt{n}}{n}\sum_{i=1}^{n} X_i$
$\mathbb{E}[\sqrt{n}\bar{X}] = \dfrac{\sqrt{n}}{n}\mathbb{E}[\sum_{i=1}^{n} X_i] = \dfrac{\sqrt{n}}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = 0\ (\because X_i \sim \mathcal{N}(0, 1))$
$\text{Var}(\sqrt{n}\bar{X}) = n * \text{Var}(\dfrac{1}{n}\sum_{i=1}^{n} X_i) = \dfrac{1}{n} * \text{Var}(\sum_{i=1}^{n} X_i) = \dfrac{n}{n} = 1$
Hence, it satisfies

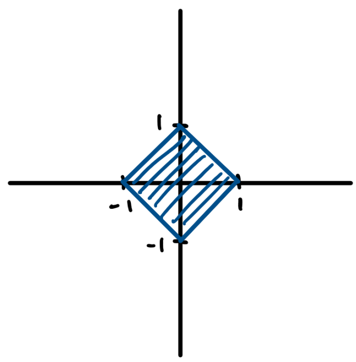
$$\sqrt{n}\bar{X} \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0, 1)$$

6 Linear algebra

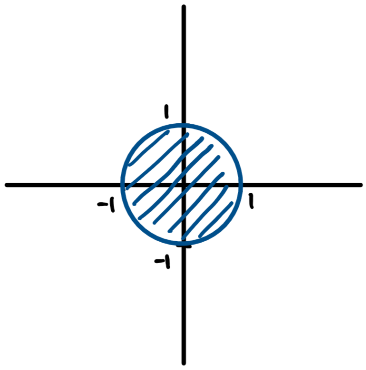
6.1 Norms [5 pts]

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

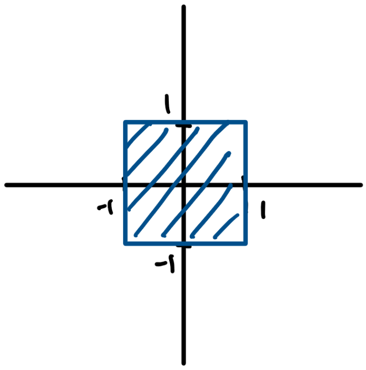
1. $||\mathbf{x}||_1 \leq 1$ (Recall that $||\mathbf{x}||_1 = \sum_i |x_i|$)



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)
   $M^T M - \lambda I = \begin{pmatrix} 25-\lambda & 0 & 0 \\ 0 & 49-\lambda & 0 \\ 0 & 0 & 9-\lambda \end{pmatrix}$,
   Then $(25-\lambda)(49-\lambda)(9-\lambda) = 0$, where the greatest eigenvalue is 49,
   $\therefore$ the norm is just $\sqrt{49} = 7$

5. $||M||_F$ (Frobenius norm)
   $\sqrt{5^2 + 7^2 + 3^2} = \sqrt{83}$

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_0 denote the origin and d be the smallest Euclidian distance.
   Then $d = ||proj_w(x_0 - x)|| = ||\dfrac{w^T(x_0 - x)}{w^T w} * w|| = |w^T x_0 - w^T x|\dfrac{||w||_2}{||w||_2^2} = \dfrac{|w^T x_0 - w^T x|}{||w||_2}$
   From the question, we know that $w^T x = -b$ and the dot product of the origin x_0 and any w will always be $= 0$, hence we can rewrite the equation as $\dfrac{|b|}{||w||_2}$, $\therefore QED$

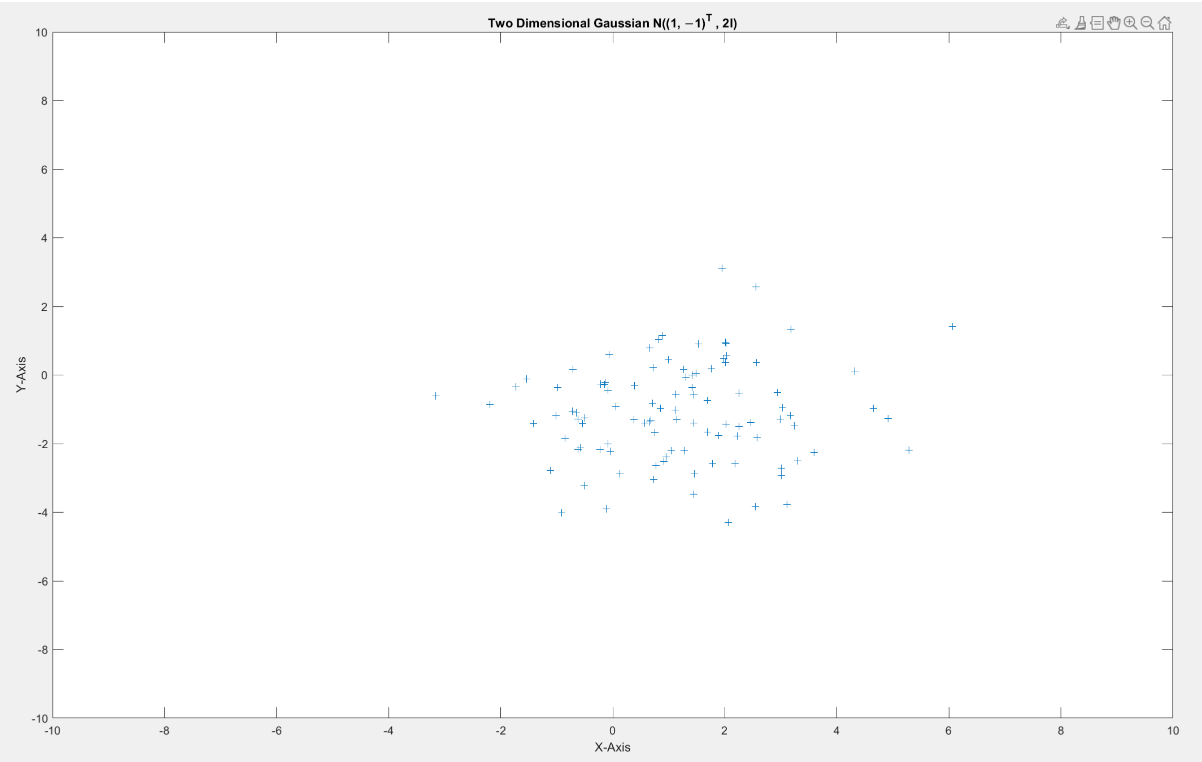
2. The Euclidean distance between two parallel hyperplane $\mathbf{w}^T\mathbf{x} + b_1 = 0$ and $\mathbf{w}^T\mathbf{x} + b_2 = 0$ is $\frac{|b_1 - b_2|}{||\mathbf{w}||_2}$ (Hint: you can use the result from the last question to help you prove this one).

   From (1), we have $\dfrac{|w^T x_0 - w^T x|}{||w||_2}$.
   If we replace x_0 to some x on the parallel hyperplane ($w^T x + b_2 = 0$), then $w^T x = -b_2$ instead of 0. Going back to the equation from (1), we will get $\dfrac{|b_1 - b_2|}{||w||_2}$

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



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