HOMEWORK 1

Jackson Hellmers
9075444662

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

   $\begin{pmatrix} 2*3 - 7*1 & 2*2 - 5*1 \end{pmatrix} * \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix}$

2. Is X invertible? If so, give the inverse, and if no, explain why not.
   Yes it is as the determinant is nonzero

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

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?

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

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 $p(x = 1) = 0.6$?

   $0.6 * 0.6 * 0.6 * 0.4 * 0.6 = 0.0518$

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.
   Probability equal to $p^4(1 - p)$
   With calculus we find it has a max when $p = 0.8$
   This yields $0.8 * 0.8 * 0.8 * 0.2 * 0.8 = 0.0819$

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 $P(A = 0 | B = 1)$?
$$\frac{0.1}{0.1 + 0.5} = 0.167$$

(b) What is $P(A = 1 \vee B = 1)$?
$0.1 + 0.1 + 0.5 = 0.7$

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

   We know that $f(n) \in O(g(n))$ if $lim_{n \to \infty} \dfrac{f(n)}{g(n)} \in [0, \infty)$

   $lim_{x \to \infty} \dfrac{ln(n)}{\log_2(n)} = ln(2) \in [0, \infty)$ Therefore both.

2. $f(n) = \log_2 \log_2(n)$, $g(n) = \log_2(n)$.
   Again using limits we find.
   $lim_{n \to \infty} \dfrac{\log_2(\log_2(n))}{\log_2(n)} = 0$
   So $f(n) = O(g(n))$

3. $f(n) = n!$, $g(n) = 2^n$.
   $lim_{n \to \infty} \dfrac{2^n}{n!} = 0$
   So $g(n) = O(f(n))$

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. $P(A|B)P(B) = P(B|A)P(A)$

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. $P(B|A) + P(B|A^c) = P(B)$

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

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 $x \in \{0, \ldots, n\}$; 0 otherwise

(a) Gamma (j)

(b) Multinomial (i)

(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 $x_i \in \{0, \ldots, n\}$ 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?
       $npq = np(1-p)$

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) $\text{Var}(5X)$
       $5^2 * 1 = 25$

   (c) $\text{Var}(X + 5)$
       1 - Shift does not change variance.

5.4   Mutual and Conditional Independence [12 pts]

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

   $E[XY] = \sum_{i=1}^n \sum_{j=1}^k x_i y_j p_{ij} =$
   $\sum_{i=1}^n x_i p_i \sum_{j=1}^k y_j p_j = E[X]E[Y]$
   This comes from us knowing that $E[XY] = E[X]E[Y]$ for independent RVs.

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

   We are told $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
   We also know for independent RVs X, Y that $Cov(X, Y) = 0$
   This comes from $Cov(X, Y) = E[XY] - E[X]E[Y]$
   So $Var(X + Y) = Var(X) + Var(Y) + 0 = Var(X) + Var(Y)$

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, 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, we now know that the result of the second die cannot be an even value.

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

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

$\tilde{X} = \frac{1}{n}\sum_{i=1}^{n} \sim \mathcal{N}(0, \frac{1}{n})$
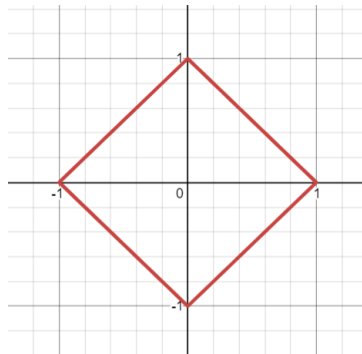We know if $Var(X) = 1$ then $Var(nX) = n^2$ so
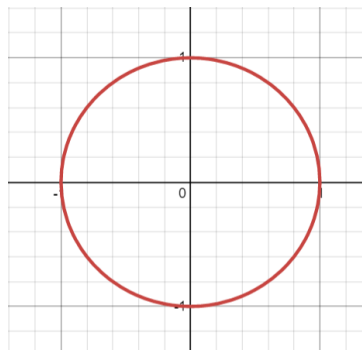$\sqrt{n}\tilde{X} \sim \mathcal{N}(0, \frac{\sqrt{n}^2}{n}) = \mathcal{N}(0,1)$

6 Linear algebra

6.1 Norms [5 pts]

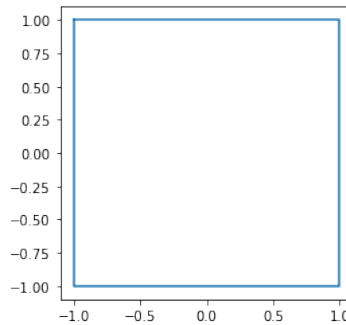Draw the regions corresponding to vectors $\mathbf{x} \in \mathbb{R}^2$ with the following norms:

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)

   $\sigma_{max} = ||M||_2 = 7$

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

   $\sqrt{5^2 + 7^2 + 3^2} = 9.11$
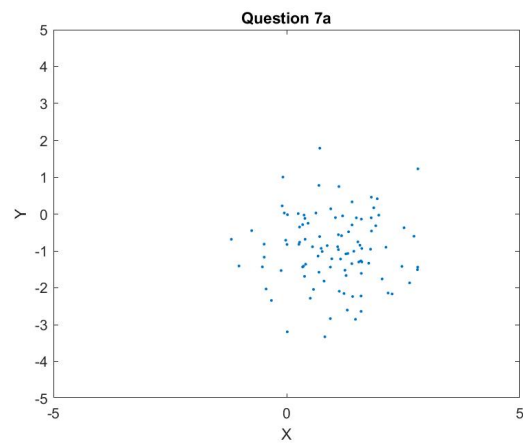
6.2   Geometry [10 pts]

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 $\mathbf{w} \neq 0$.

   $w^T x = w \cdot x = ||w||_2 ||x||_2 \cos(\theta)$ - definition of dot product.

   Where $||x||_2 \cos(\theta)$ is the projection of x onto w.

   So $\dfrac{w^T x}{||w||_2} = \dfrac{|b|}{||w||_2} = ||x||_2 \cos(\theta)$ which as we already know is the projection of x onto w.

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).

   In this problem our distance will be equal to $proj_w(X_1 - X_2)$

   Given what we know from the last question.

   $proj_w(X_1) = \dfrac{-b_1}{||w||_2}$ and $proj_w(X_2) = \dfrac{-b_2}{||w||_2}$

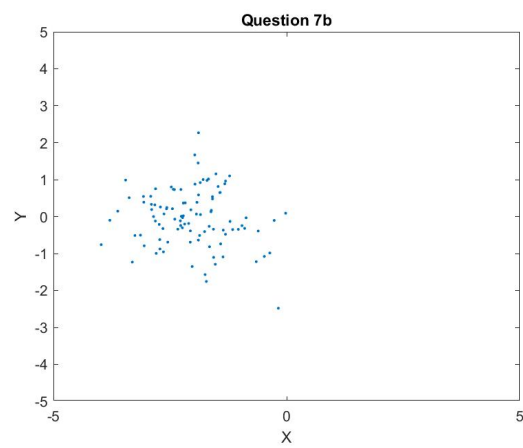   So $proj_w(X_1 - X_2) = \dfrac{-b_1 + b_2}{||w||_2} = \dfrac{|b_1 - b_2|}{||w||_2}$

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 $\mathbb{R}^{2\times2}$.

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

Graph for 7.1

Graph for 7.2