>>XX<<
>>XXX<<

Minimum Background Test [80 pts]

1 Vectors and Matrices [20 pts]

Consider the matrix X and the vectors $\mathbf{y}$ and $\mathbf{z}$ below:

$$X = \begin{pmatrix} 9 & 8 \\ 7 & 6 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} 9 \\ 8 \end{pmatrix} \qquad \mathbf{z} = \begin{pmatrix} 7 \\ 6 \end{pmatrix}$$

1. What is the inner product of the vectors $\mathbf{y}$ and $\mathbf{z}$? (this is also sometimes called the *dot product*, and is sometimes written as $\mathbf{y}^T\mathbf{z}$)

$$\begin{aligned} y^Tz &= \begin{pmatrix} 9 & 8 \end{pmatrix}\begin{pmatrix} 7 \\ 6 \end{pmatrix} \\ &= 9 \times 7 + 8 \times 6 \\ &= 104 \end{aligned}$$

2. What is the product $X\mathbf{y}$?

$$\begin{aligned} Xy &= \begin{pmatrix} 9 & 8 \\ 7 & 6 \end{pmatrix}\begin{pmatrix} 9 \\ 8 \end{pmatrix} \\ &= \begin{pmatrix} 9 \times 9 + 8 \times 8 \\ 7 \times 9 + 6 \times 8 \end{pmatrix} \\ &= \begin{pmatrix} 145 \\ 111 \end{pmatrix} \end{aligned}$$

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

$$Xi = \begin{pmatrix} -3 & 4 \\ 3.5 & -4.5 \end{pmatrix}$$

4. What is the rank of X?

$$X \Rightarrow Xref = \begin{pmatrix} 1 & 8 \\ 0 & -3 \end{pmatrix}$$

Because the $Xref$ has two none-zero rows, the rank of X is 2.

2 Calculus [20 pts]

1. If $y = 4x^3 - x^2 + 7$ then what is the derivative of y with respect to x?

$$\frac{\partial y}{\partial x} = 12x^2 - 2x$$

2. If $y = \tan(z)x^{6z} - \ln(\frac{7x+z}{x^4})$, what is the partial derivative of y with respect to x?

$$\begin{aligned}
\frac{\partial y}{\partial x} &= \tan(z) \times 6z \times x^{6z-1} - \frac{\partial \frac{7x+z}{x^4}}{\partial x} \times \frac{x^4}{7x+z} \\
&= 6z\tan(z)x^{6z-1} - \frac{7}{7x+z} + \frac{4}{x}
\end{aligned}$$

3 Probability and Statistics [20 pts]

Consider a sample of data $S = \{0, 1, 1, 0, 0, 1, 1\}$ created by flipping a coin x seven times, where 0 denotes that the coin turned up heads and 1 denotes that it turned up tails.

1. What is the sample mean for this data?

$$\begin{aligned}
\bar{x} &= \frac{\sum_{i=1}^{N} x_i}{N} \\
&= (0 + 1 + 1 + 0 + 0 + 1 + 1) \div 7 \\
&= \frac{4}{7} \approx 0.571
\end{aligned}$$

2. What is the sample variance for this data?

$$\begin{aligned}
\sigma^2 &= \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N} \\
&= (3 \times (\frac{3}{7})^2 + 4 \times (\frac{4}{7})^2) \div 7 \\
&= \frac{12}{49} \approx 0.245
\end{aligned}$$

3. What is the probability of observing this data, assuming it was generated by flipping a biased coin with $p(x = 1) = 0.7, p(x = 0) = 0.3$.

$$\begin{aligned}
p &= C_7^3 \times 0.3^3 \times 0.7^4 \\
&\approx 0.227
\end{aligned}$$

4. Note that the probability of this data sample would be greater if the value of $p(x = 1)$ was not 0.7, but instead some other value. What is the value that maximizes the probability of the sample S? Please justify your answer.

Let

$$k = p(x = 1)$$

,
then

$$p(x = 0) = 1 - k$$

.
The probability of the sample S is:

$$p = C_7^3 \times k^4 \times (1 - k)^3$$

the derivative of p with respect to k is:

$$
\begin{aligned}
\frac{\partial p}{\partial k} &= 35 \times (4 \times k^3(1-k)^3 - 3 \times (1-k)^2 k^4) \\
&= 35 \times (4 - 7k) \times (1 - k)^2 \times k^3
\end{aligned}
$$

So when $p(x = 1) = \frac{4}{7} \approx 0.571$, the probability of the sample S can be maximazed.

5. Consider the following joint probability table where both A and B are binary random variables:

A	B	$P(A, B)$
0	0	0.1
0	1	0.4
1	0	0.2
1	1	0.3

   (a) What is $P(A = 0, B = 0)$?
       $P(A = 0, B = 0) = 0.1$

   (b) What is $P(A = 1)$?
       $P(A = 1) = 0.2 + 0.3 = 0.5$

   (c) What is $P(A = 0 | B = 1)$?
       $P(A = 0 | B = 1) = 0.4 \div (0.4 + 0.3) = \frac{4}{7} \approx 0.571$

   (d) What is $P(A = 0 \vee B = 0)$?
       $P(A = 0 \vee B = 0) = 1$

4 Big-O Notation [20 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. $f(n) = \frac{n}{2}$, $g(n) = \log_2(n)$.
   $g(n) = O(f(n))$. $f(n)$ grows faster than $g(n)$ when n becomes very large.

2. $f(n) = \ln(n)$, $g(n) = \log_2(n)$
   Both. $g(n)$ and $f(n)$ both equal to the same order when n gets very large.

3. $f(n) = n^{100}$, $g(n) = 100^n$.
   $f(n) = O(g(n))$. $g(n)$ grows faster than $f(n)$ when n becomes very large.

3

Medium Background Test [20 pts]

5 Algorithm [5 pts]

Divide and Conquer: Assume that you are given a sorted array with n integers in the range $[-10, +10]$. Note that some integer values may appear multiple times in the array. Additionally, you are told that somewhere in the array the integer 0 appears exactly once. Provide an algorithm to locate the 0 which runs in $O(\log(n))$. Explain your algorithm in words, describe why the algorithm is correct, and justify its running time.

This problem can be solved by Binary Search. In the while loop, every iteration will cover half of the array, and shorten the searching space into half of the current length. So the big O complexity will be $O(log(n))$. The fake code is below:

Input: $target = 0$, $nums$ is a sorted array (accending)
Output: index of 0
  $start \leftarrow 0$
  $end \leftarrow len(nums) - 1$
  while $start + 1 < end$ **do**
    $mid \leftarrow start + (end - start)/2$
    if $nums[mid] < 0$ **then**
      $start \leftarrow mid$
    else if $nums[mid] > 0$ **then**
      $end \leftarrow mid$
    else
      return mid
    end if
  end while
  if $nums[start] == 0$ **then**
    return $start$
  else
    return end
  end if

---

6 Probability and Random Variables [5 pts]

6.1 Probability

State true or false. Here Ω denotes the sample space and A^c denotes the complement of the event A.

1. For any $A, B \subseteq \Omega$, $P(A|B)P(B) = P(B|A)P(A)$.
   true

2. For any $A, B \subseteq \Omega$, $P(A \cup B) = P(A) + P(B) - P(A|B)$.
   false

3. For any $A, B, C \subseteq \Omega$ such that $P(B \cup C) > 0$, $\frac{P(A \cup B \cup C)}{P(B \cup C)} \geq P(A|B \cup C)P(B \cup C)$.
   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. For any n events $\{A_i\}_{i=1}^{n}$, if $P(\bigcap_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$, then $\{A_i\}_{i=1}^{n}$ are mutually independent.
   true

6.2 Discrete and Continuous Distributions

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) Laplace (h)

(b) Multinomial (i)

(c) Poisson (l)

(d) Dirichlet (k)

(e) Gamma (j)

(h) $f(x; b, \mu) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$

(i) $f(\boldsymbol{x}; n, \boldsymbol{\alpha}) = \frac{n!}{\prod_{i=1}^k x_i!}\Pi_{i=1}^k \alpha_i^{x_i}$ 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

6.3 Mean and Variance

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?
   $\mu_x = np$

   (b) What is the variance of the random variable?
   $\sigma_x^2 = np(1 - p)$

2. Let X be a random variable and $\mathbb{E}[X] = 1, \text{Var}(X) = 1$. Compute the following values:

   (a) $\mathbb{E}[3X]$
   $\mathbb{E}[3X] = 3 \times \mathbb{E}[X] = 3$

   (b) $\text{Var}(3X)$
   $\text{Var}(3X) = 3^2 \times \text{Var}(X) = 9$

   (c) $\text{Var}(X + 3)$
   $\text{Var}(X + 3) = \text{Var}(X) = 1$

6.4 Mutual and Conditional Independence

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

   For independent random variables X and Y,

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_i^N \sum_i^N x_i y_i f(xi, yi) \\
&= \sum_i^N \sum_i^N x_i y_i f(x_i) f(y_i) \\
&= \sum_i^N x_i f(x_i) \sum_i^N y_i f(y_i) \\
&= \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

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

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

Since X and Y are independent random variables, so

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$$

Then,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

3. 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.

   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?

   Yes.

6.5   Law of Large Numbers and the Central Limit Theorem

Provide one line justifications.

1. Suppose we simultaneously flip two independent fair coins (i.e., the probability of heads is $1/2$ for each coin) and record the result. After 40,000 repetitions, the number of times the result was two heads is close to 10,000. (Hint: calculate how close.)

   The probability that the result is two heads is $0.5^2 = 0.25$. So after 40,000 repetitions, the result that was two heads should be close to $0.25 \times 40,000 = 10,000$.

2. 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

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

   According to central limit theorem, let $S_n = \sum_{i=1}^{n} X_i$, there is $\lim_{n\to\infty} P(\frac{S_n - \mu n}{\sigma \sqrt{n}}) \sim \mathcal{N}(0, 1)$, $\mu = 0$, $\sigma = 1$
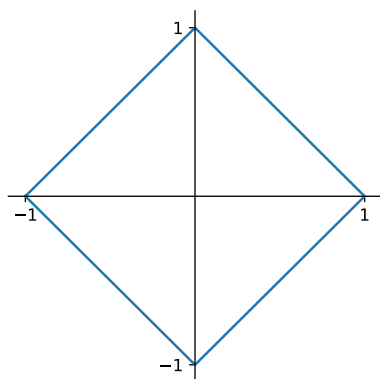   Then $\lim_{n\to\infty} P(\frac{S_n}{\sqrt{n}}) \sim \mathcal{N}(0, 1)$
   $\sqrt{n}\bar{X} = \frac{1}{n} \times \sqrt{n} \times S_n = \frac{S_n}{\sqrt{n}} \sim \mathcal{N}(0, 1)$
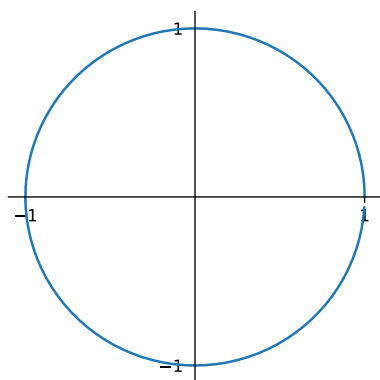
7   Linear algebra [5 pts]

7.1   Norm-enclature

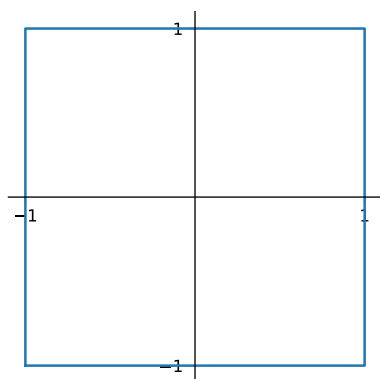Draw the regions corresponding to vectors $\mathbf{x} \in \mathbb{R}^2$ with the following norms:

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

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

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

7.2 Geometry

Prove that these are true or false. 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 $\frac{|b|}{||\mathbf{w}||_2}$.

   true.

   The distance from point P to hyperplane $\mathbf{w}^T\mathbf{x} + b = 0$ is along a line parallel to the normal vector, which is

   $$\vec{N} = \frac{\mathbf{w}}{||\mathbf{w}||_2}$$

   for point P, the distance between P and any point Q on this hyperplane can be presented as

   $$\mathbf{v} = |x_p - x_q|$$

   the shortest distance is the projection of $\mathbf{v}$ on the normal vector $\vec{N}$.

   $$D = \frac{|\mathbf{w}^T(x_p - x_q)|}{||\mathbf{w}||_2}$$

   Because Q is on the plane, $\mathbf{w}^T\mathbf{x_q} + b = 0$, so

   $$D = \frac{|\mathbf{w}^T x_p + b|}{||\mathbf{w}||_2}$$

   Because x_p is the origin, the shortest path is

   $$D = \frac{|b|}{||\mathbf{w}||_2}$$

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).

   true.

   for hyperplane 1, the shortest path between the origin with the plane is

   $$D_1 = \frac{|b_1|}{||\mathbf{w}||_2}$$

   for hyperplane 2, the shortest path between the origin with the plane is

   $$D_2 = \frac{|b_2|}{||\mathbf{w}||_2}$$

   Because these two hyperplanes are parellel, and the shortest distance between the origin and hyperplane are both vertical. So the distance between these two hyperplanes is

   $$\begin{aligned} D &= \frac{|b_1|}{||\mathbf{w}||_2} - \frac{|b_2|}{||\mathbf{w}||_2} \\ &= \frac{|b_1 - b_2|}{||\mathbf{w}||_2} \end{aligned}$$

# 8 Programming Skills - Matlab [5pts]

Sampling from a distribution. For each question, submit a scatter plot (you will have 5 plots in total). Make sure the axes for all plots have the same limits. (Hint: You can save a Matlab figure as a pdf, and then use includegraphics to include the pdf in your latex file.)
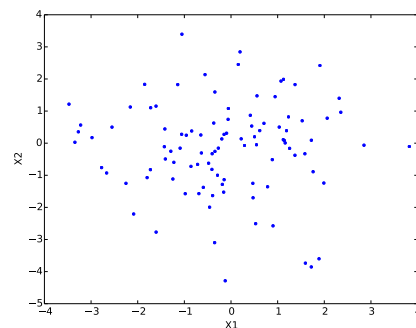
1. Draw 100 samples $\mathbf{x} = [x_1, x_2]^T$ from a 2-dimensional Gaussian distribution with mean $(0, 0)^T$ and identity covariance matrix, i.e., $p(\mathbf{x}) = \frac{1}{2\pi} \exp\left(-\frac{||\mathbf{x}||^2}{2}\right)$, and make a scatter plot (x_1 vs. x_2). For each question below, make each change separately to this distribution.
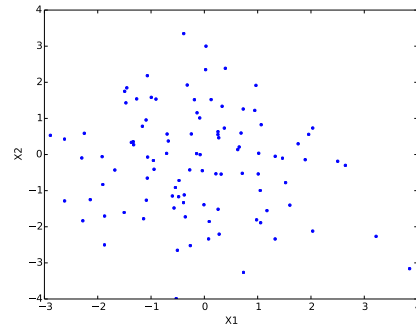
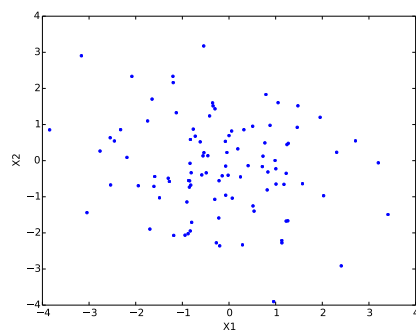2. Make a scatter plot with a changed mean of $(1, -1)^T$.
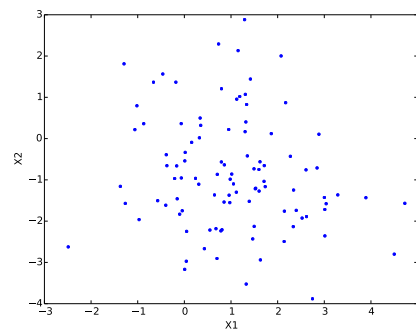
3. Make a scatter plot with a changed covariance matrix of $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.

4. Make a scatter plot with a changed covariance matrix of $\begin{pmatrix} 2 & 0.2 \\ 0.2 & 2 \end{pmatrix}$.

5. Make a scatter plot with a changed covariance matrix of $\begin{pmatrix} 2 & -0.2 \\ -0.2 & 2 \end{pmatrix}$.