HOMEWORK 1:
BACKGROUND TEST

>>NAME HERE<<
>>ID HERE<<

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 $\mathbf{y}^T\mathbf{z}$)
   111.(5pts)

2. What is the product $X\mathbf{y}$?
   $\begin{pmatrix} 145 \\ 111 \end{pmatrix}$. (5pts)

3. Is X invertible? If so, give the inverse, and if no, explain why not.
   X is invertible. The inverse matrix is
   $\begin{pmatrix} -3 & 4 \\ 3.5 & -4.5 \end{pmatrix}$. (5pts)

4. What is the rank of X?
   Rank(X) = 2. (5pts)

2 Calculus [20 pts]

1. If $y = 4x^3 - x^2 + 7$ then what is the derivative of y with respect to x?
   $\frac{dy}{dx} = 12x^2 - 2x$. (10pts)

2. If $y = \tan(z)x^{6z} - \ln(\frac{7x+z}{x^4})$, what is the partial derivative of y with respect to x?
   $\frac{dy}{dx} = 6ztan(z)x^{6z-1} - \frac{7}{7x+z} + \frac{4}{x}$.
   (10pts)

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?
   $\frac{4}{7}$ or 0.5714285714285714. (4pts)

2. What is the sample variance for this data?
   Full credits would be given for those two answers:

   $12/49$ or 0.24489795918367346(divide by n);

2/7 or 0.2857142857142857(divide by n-1)

(4pts)

3. What is the probability of observing this data, assuming it was generated by flipping a biased coin with $p(x = 1) = 0.7, p(x = 0) = 0.3$.
$(0.3)^3(0.7)^4 = 0.006482699999999975$. (4pts)

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.
Probability of the given sample
$P = p^4 * (1 - p)^3$, for P to be max, $\frac{dP}{dp} = 0$. $p = \frac{4}{7} = 0.5714285714285714$. (4pts)

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. (1pt)

(b) What is $P(A = 1)$?
$P(A = 1) = 0.5$. (1pt)

(c) What is $P(A = 0|B = 1)$?
$P(A = 0|B = 1) = \frac{4}{7}$. (1pt)

(d) What is $P(A = 0 \vee B = 0)$?
$P(A = 0 \vee B = 0) = 0.7$. (1pt)

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) = O(g(n)) iff exists n0, c such that for all $n > n0, |f(n)| < cg(n)$. Therefore, we need to find n0 and c.

For f(n) = n/2, g(n) = log(n), we can design c = 2 to simplify $log(n) < cn/2 = n$ for all $n > 0$.

Full credits would also be given if using limit of $n \to \infty$ to justify.

(6pts)

2. $f(n) = \ln(n), g(n) = \log_2(n)$.
Similar to the above proof, g(n) = O(f(n)) and f(n) = O(g(n)).

(8pts; 4 for g(n) = O(f(n)) and 4 for f(n) = O(g(n)))

3. $f(n) = n^{100}, g(n) = 100^n$.
f(n) = O(g(n)).

(6pts)

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.

We can use binary search: keep three indexes, low, mid and high.
At first low = 0, high = n-1, mid = [(low+high)/2]. Every time we compare array[mid] with zero, if array[mid]>0 then 0 is definitely in the left part of the array, so let high = mid-1 and mid = [(low+high)/2]; else let low = mid+1 and mid = [(low+high)/2]; until we find array[mid] == 0, or low >= high.
Since the array is sorted, every loop we narrow the possible indexes that 0 might be at least by half, so the algorithm will find 0 in log(n).
(Binary search 3pts; explanation 2pts.)

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. (0.2; the same with below.)
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, $|\boldsymbol{x}| = k$.

(a) Laplace h (0.2 for each)

(b) Multinomial i

(c) Poisson l

(d) Dirichlet k

(e) Gamma j

(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

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?
   np. (0.2 for each)

   (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) $\mathbb{E}[3X]$
   3. (0.2 for each)

   (b) $\text{Var}(3X)$
   9.

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

6.4 Mutual and Conditional Independence

1. If X and Y are independent random variables, show that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
   $E[XY] = \int_{-\infty}^{\infty} p(X,Y)dxdy = \int_{-\infty}^{\infty} p(X)p(Y)dxdy = \int_{-\infty}^{\infty} p(X)dx \int_{-\infty}^{\infty} p(Y)dy = E[X]E[Y]$.
   (0.3; full credits would be given if using independent variables P(AB) = P(A)P(B).

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

   $Cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y] = 0$. So $\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y) = \text{Var}(X) + \text{Var}(Y)$.
   (0.3; full credits would be given if using E[XY] = E[X]E[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, since they are independent.

   (0.2)

   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. 1st die result is 1 and the sum of 2 results is even. So possible cases are only 1,3,5. (0.2)

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.)

   Let X_i denote if result is two heads in the i-th experiment. $P(X_i = 1) = \frac{1}{4}$, $P(X_i = 0) = \frac{3}{4}, E(X_i) = \frac{1}{4}$. X_i is binomial.

   Let $\bar{X} = \frac{1}{n}\Sigma_{i=1}^{n} X_i$. Using Chebyshev's inequality and binomial quality, $P(|\bar{X} - E(X_i)| > t) < \frac{\sigma^2}{nt^2}$, where $\sigma = var(X_i)$. Let n = 40000, we can get how close the result is from the mean value.

   (0.5; For full credits just explaining the convergence of the sum of independent variables is OK.)

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

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

E$[\sqrt{n}\bar{X}] = \frac{1}{\sqrt{n}}E[\Sigma_i X_i] = 0.$

Var$[\sqrt{n}\bar{X}] = \frac{1}{n}Var[\Sigma_i X_i] = 1.$

(0.5; full credits would be given as long as describing the central limit theorem.)

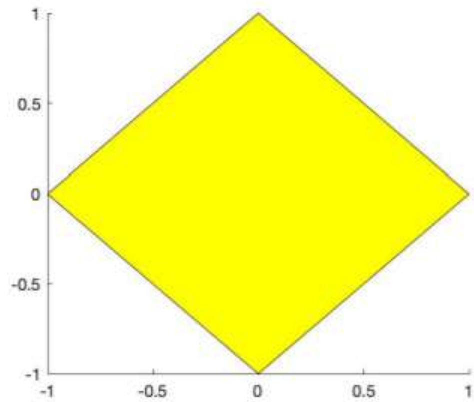
7 Linear algebra [5 pts]

7.1 Norm-enclature

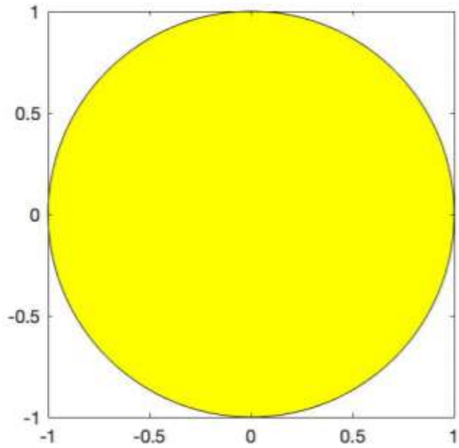
Draw the regions corresponding to vectors $\mathbf{x} \in \mathbb{R}^2$ with the following norms:

1. $||\mathbf{x}||_1 \leq 1$ (Recall that $||\mathbf{x}||_1 = \sum_i |x_i|$)
2. $||\mathbf{x}||_2 \leq 1$ (Recall that $||\mathbf{x}||_2 = \sqrt{\sum_i x_i^2}$)
3. $||\mathbf{x}||_\infty \leq 1$ (Recall that $||\mathbf{x}||_\infty = \max_i |x_i|$)

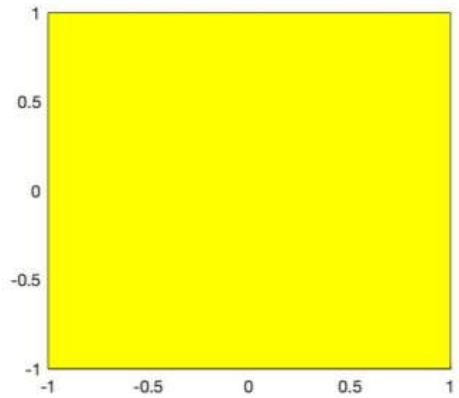
(1 pt for each.) 1.



2.



3.



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 $\frac{|b|}{||\mathbf{w}||_2}$.
   Assume that x is in the hyperplane. We have $||w||_2||x||_2 \geq ||w^Tx||_2 = |b|$ from Cauchy-Schwarz inequality, and only if $w = Cx$ (C is a constant) does the equality hold. So the smallest Eucilidean distance is $\frac{|b|}{||\mathbf{w}||_2}$.
   (1pt; using projection to get the same result is also OK)

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).

   Assume that $\mathbf{w}^\top\mathbf{x_1} + b_1 = 0$ and $\mathbf{w}^\top\mathbf{x_2} + b_2 = 0$. Still using the Cauchy-Schwarz inequality we have $||w||_2||x_1-x_2||_2 \geq ||w^T(x_1-x_2)||_2 = |b_1-b_2|$. So the Euclidean distance between two parallel hyperplane is $\frac{|b_1-b_2|}{||\mathbf{w}||_2}$.

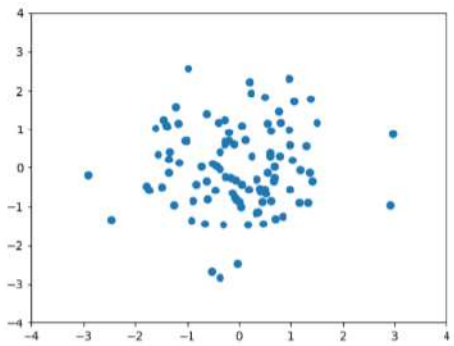
   (1pt; using projection to get the same result is also OK)

8 Programming Skills [5pts]

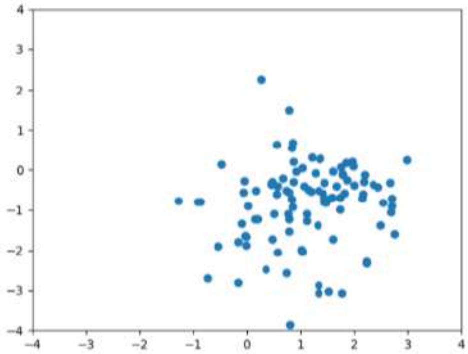
Sampling from a distribution. For each question, submit a scatter plot (you will have 5 plots in total). Make sure the axes for all plots have the same limits. (Hint: You can save the figure as a pdf, and then use includegraphics to include the pdf in your latex file. Suggest to use Python or Matlab.)

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.

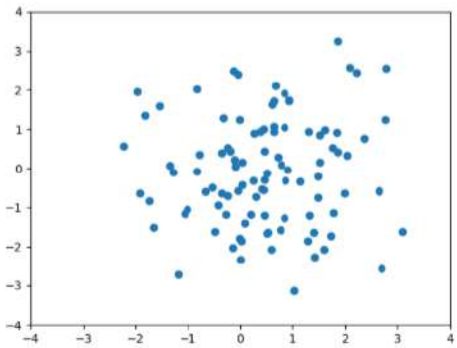
   (1pt for each; full credits would be given depending on the rough mean position and shape.)



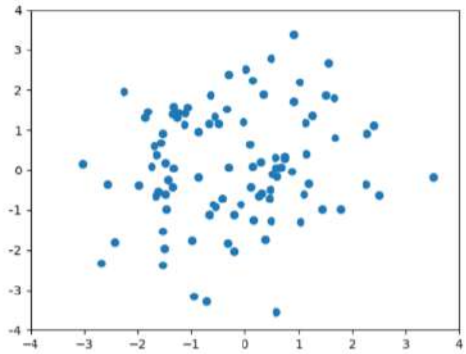
2. Make a scatter plot with a changed mean of $(1,-1)^T$.



3. Make a scatter plot with a changed covariance matrix of $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.
   Solution figure goes here.



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 $\begin{pmatrix} 2 & -0.2 \\ -0.2 & 2 \end{pmatrix}$.