HOMEWORK 1:
BACKGROUND TEST

>>ARPIT JAIN<<
>>ajain74<<

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

   if $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, and $\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$
   , then $\mathbf{y}.\mathbf{z} = y1.z1 + y2.z2$
   $\implies \mathbf{y}.\mathbf{z} = (9*7) + (8*6)$
   $\implies \mathbf{y}.\mathbf{z} = 63 + 48$

   $$\implies \mathbf{y}.\mathbf{z} = 111$$

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

   if $X = \begin{pmatrix} x1 & x2 \\ x3 & x4 \end{pmatrix}$, and $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$
   , then $X\mathbf{y} = \begin{pmatrix} x1.y1 + x2.y2 \\ x3.y1 + x4.y2 \end{pmatrix}$
   $\implies X\mathbf{y} = \begin{pmatrix} (9*9) + (8*8) \\ (7*9) + (6*8) \end{pmatrix}$
   $\implies X\mathbf{y} = \begin{pmatrix} 81 + 64 \\ 63 + 48 \end{pmatrix}$

   $$\implies X\mathbf{y} = \begin{pmatrix} 145 \\ 111 \end{pmatrix}$$

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

   As we know, a matrix X is invertible if $det(x) \neq 0$
   now, $det(x) = x1.x4 - x2.x3$
   $\implies det(x) = x1.x4 - x2.x3$
   $\implies det(x) = (9*6) - (8*7)$
   $\implies det(x) = 54 - 56$
   $\implies det(x) = -2 \neq 0$
   hence, **X is invertible**

   now, $XX^{-1} = I$
   say, $X^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

then, $\begin{pmatrix} 9 & 8 \\ 7 & 6 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$$9a + 8c = 1 \text{ .... } (1)$$
$$9b + 8d = 0 \text{ .... } (2)$$
$$7a + 6c = 0 \text{ .... } (3)$$
$$7b + 6d = 1 \text{ .... } (4)$$

using (2), we get

$$d = \frac{-9b}{8} \text{ .... } (5)$$

using (3), we get

$$c = \frac{-7a}{6} \text{ .... } (6)$$

substituting (5) in (4), we get

$7b + 6(\frac{-9b}{8}) = 1$
$\implies b = 4$
$\implies d = \frac{-9}{2}$

substituting (6) in (1), we get $9a + 8(\frac{-7a}{6}) = 1$
$\implies a = -3$
$\implies c = \frac{7}{2}$

$$\implies \mathbf{X^{-1}} = \begin{pmatrix} -3 & 4 \\ \frac{7}{2} & \frac{-9}{2} \end{pmatrix} = \begin{pmatrix} -3 & 4 \\ 3.5 & -4.5 \end{pmatrix}$$

4. What is the rank of X?
given, $X = \begin{pmatrix} 9 & 8 \\ 7 & 6 \end{pmatrix}$
now, to find rank(X), we will transform X to its row echelon form

$\implies \begin{pmatrix} 9 & 8 \\ 7 & 6 \end{pmatrix} -> \begin{pmatrix} 1 & x \\ 0 & y \end{pmatrix}$

replacing $r_1 -> \frac{r_1}{9}$
$\implies \begin{pmatrix} 9 & 8 \\ 7 & 6 \end{pmatrix} -> \begin{pmatrix} 1 & \frac{8}{9} \\ 7 & 6 \end{pmatrix}$

replacing $r_2 -> r_2 - 7r_1$
$\implies \begin{pmatrix} 1 & \frac{8}{9} \\ 7 & 6 \end{pmatrix} -> \begin{pmatrix} 1 & \frac{8}{9} \\ 0 & \frac{-2}{9} \end{pmatrix}$

now we have 2 independent rows in the row echelon form of X
hence,

$$\mathbf{rank(X) = 2}$$

2 Calculus [20 pts]

1. If $y = 4x^3 - x^2 + 7$ then what is the derivative of y with respect to x?
given, $y = 4x^3 - x^2 + 7$
$\implies \frac{dy}{dx} = \frac{d}{dx}(4x^3 - x^2 + 7)$
$\implies \frac{dy}{dx} = \frac{d}{dx}4x^3 + \frac{d}{dx}(-x^2) + \frac{d}{dx}7$
$\implies \frac{dy}{dx} = (3).4x^{4-1} - 2x^{2-1}) + 0$

$= (3).4x^{4-1} - 2x^{2-1}) + 0$

$$\implies \frac{dy}{dx} = 12x^2 - 2x \; or$$

$$\implies \frac{dy}{dx} = 2x(6x-1)$$

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

using chain rule,
$\implies \quad = (\tan(z))(6z)(x^{6z-1}) - \frac{\partial}{\partial(7x+z)}(\ln(7x+z))\frac{\partial}{\partial x}(7x+z) + \frac{\partial}{\partial x^4}(\ln(x^4))\frac{\partial}{\partial x}(x^4)$
$\implies \quad = (\tan(z))(6z)(x^{6z-1}) - \frac{1}{7x+z}(7) + \frac{1}{x^4}(4x^3)$

$$\implies \frac{\partial y}{\partial x} = (\tan(z))6zx^{6z-1} - \frac{7}{7x+z} + \frac{4}{x}$$

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?

mean = $\frac{sum of all observations}{total number of observations}$
$\implies \quad = \frac{0+1+1+0+0+1+1}{7}$

$$\implies mean = \frac{4}{7}$$

2. What is the sample variance for this data?

variance = $\frac{1}{N}\sum_{i=1}^{N}(x_i - mean)^2$
$\implies \quad = \frac{1}{7}[(0-\frac{4}{7})^2 + (1-\frac{4}{7})^2 + (1-\frac{4}{7})^2 + (0-\frac{4}{7})^2 + (0-\frac{4}{7})^2 + (1-\frac{4}{7})^2 + (1-\frac{4}{7})^2]$
$\implies \quad = \frac{1}{7}[3(0-\frac{4}{7})^2 + 4(1-\frac{4}{7})^2]$
$\implies \quad = \frac{1}{7}[3(\frac{16}{49}) + 4(\frac{3}{7})^2]$
$\implies \quad = \frac{1}{7}[\frac{48}{49} + 4(\frac{9}{49})]$
$\implies \quad = \frac{1}{7}[\frac{48}{49} + \frac{36}{49}]$
$\implies \quad = \frac{1}{7}(\frac{48+36}{49})$
$\implies \quad = \frac{1}{7}(\frac{84}{49})$

$$\implies variance = \frac{12}{49}$$

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

as all the events are i.i.d., hence
$P(S = \{0, 1, 1, 0, 0, 1, 1\})$
$= p(x = 0) * p(x = 1) * p(x = 1) * p(x = 0) * p(x = 0) * p(x = 1) * p(x = 1)$
$\implies \quad = (0.3)(0.7)(0.7)(0.3)(0.3)(0.7)(0.7)$
$\implies \quad = (0.3)^3(0.7)^4$

$$\implies P(S = \{0, 1, 1, 0, 0, 1, 1\}) = 0.0064827$$

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, $p(x = 1) = p$
$\implies p(x = 0) = 1 - p$
now, we need to maximize $p(x = 0)^3.p(x = 1)^4$

$\implies max((1 - p)^3.p^4)$
$\implies \frac{d}{dp}((1 - p)^3.p^4) = 0$ ...(1)
& $\frac{d^2}{dp^2}((1 - p)^3.p^4) < 0$ ...(2)

solving for (1),
$\frac{d}{dp}(p^4 - 3p^5 + 3p^6 - p^7) = 0$
$\implies 4p^3 - 15p^4 + 18p^5 - 7p^6 = 0$
$\implies p^3(4 - 15p + 18p^2 - 7p^3) = 0$
$\implies p^3(p - 1)(7p^2 - 11p + 4) = 0$
$\implies p^3(p - 1)(p - 1)(p - \frac{4}{7}) = 0$ ...(3)

solving for (2),
$\frac{d^2}{dp^2}((1 - p)^3.p^4) < 0$
$\implies \frac{d}{dp}(4p^3 - 15p^4 + 18p^5 - 7p^6) < 0$
$\implies 12p^2 - 60p^3 + 90p^4 - 42p^5 < 0$ ....(4)

using (3) and (4), $p = \frac{4}{7}$ satisfies both the conditions. hence,

$$p(x = 1) = p = \frac{4}{7}$$

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 $P(A = 1)$?
$P(A = 1) = P(A = 1, B = 0) + P(A = 1, B = 1)$
$\implies P(A = 1) = 0.2 + 0.3$

$$\implies P(A = 1) = 0.5$$

(c) What is $P(A = 0|B = 1)$?

$P(A = 0|B = 1) = \frac{P(A=0,B=1)}{P(B=1)}$
$\implies \quad = \frac{P(A=0,B=1)}{P(A=0,B=1)+P(A=1,B=1)}$
$\implies \quad = \frac{0.4}{0.4+0.3}$

$$\implies P(A = 0|B = 1) = \frac{4}{7}$$

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

$P(A = 0 \vee B = 0) = P(A = 0, B = 0) + P(A = 0, B = 1) + P(A = 1, B = 0)$
$\implies \quad = 0.1 + 0.4 + 0.2$

$$\implies P(A = 0 \vee B = 0) = 0.7$$

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. $f(n) = \frac{n}{2}$, $g(n) = \log_2(n)$.
   let us evaluate if $f(n)$ is the upper limit of $g(n)$.
   $\lim_{n\to\infty} \frac{f(n)}{g(n)} = \lim_{n\to\infty} \frac{\frac{n}{2}}{\log_2(n)}$
   $\implies \quad = \lim_{n\to\infty} \frac{n}{2} \frac{\ln(2)}{\ln(n)}$
   $\implies \quad = \lim_{n\to\infty} \frac{\ln(2)}{2} \frac{n}{\ln(n)}$

   as we now, $y = n$ increases at a much faster rate compared to $y = \log_e(n)$
   thus, $\lim_{n\to\infty} \frac{f(n)}{g(n)} = \infty$

   $\implies f(n)$ is the upper bound for $g(n)$
   $\implies g(n)$ is **NOT** the upper bound for $f(n)$

$$\implies g(n) = O(f(n))$$

2. $f(n) = \ln(n)$, $g(n) = \log_2(n)$.
   let us evaluate if $f(n)$ is the upper limit of $g(n)$.
   $\lim_{n\to\infty} \frac{f(n)}{g(n)} = \lim_{n\to\infty} \frac{\ln(n)}{\log_2(n)}$
   $\implies \quad = \lim_{n\to\infty} \frac{\ln(n)}{\frac{\ln(n)}{\ln(2)}}$
   $\implies \quad = \lim_{n\to\infty} \ln(2)$

   thus, both the functions limit each other as $n \to \infty$

$$\implies f(n) = O(g(n))$$
$$\implies g(n) = O(f(n))$$

3. $f(n) = n^{100}$, $g(n) = 100^n$.
   let us evaluate if $f(n)$ is the upper limit of $g(n)$.
   $\lim_{n\to\infty} \frac{f(n)}{g(n)} = \lim_{n\to\infty} \frac{n^{100}}{100^n}$

   As both $f(n)$ and $g(n) \to \infty$ as $n \to \infty$, hence using l'Hpital's Rule
   $\lim_{n\to\infty} \frac{f(n)}{g(n)} = \lim_{n\to\infty} \frac{f'(n)}{g'(n)}$
   $\implies = \lim_{n\to\infty} \frac{100n^{99}}{100^n . \ln(100)}$

   again, using l'Hpital's Rule
   $\implies = \lim_{n\to\infty} \frac{100.99.n^{98}}{100^n.(\ln(100))^2}$

   iteratively using l'Hpital's Rule, we get
   $\implies = \lim_{n\to\infty} \frac{100!}{100^n.(\ln(100))^{100}}$
   $\implies \lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$

   thus, $g(n)$ will be the upper bound for $f(n)$ as $n \to \infty$

$$\implies f(n) = O(g(n))$$

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 approach this problem using the binary search which will provide a time complexity of $O(\log(n))$, as there is only single occurrence of 0.

```
int binarySearch(int x, int[] A, int left, int right) {
    int m = (left + right)/2;
    if (x == A[m]) {
        return m;
    }
    if (x < A[m]){
        return binarySearch(x, A, l, m - 1)
    }else
        return binarySearch(x, A, m + 1, r);
    }
}
```

For this, let us assume the array to be A and initial index $i = \frac{n}{2}$ for the algorithm.

Now, if $A[i] < 0$, then we will search for 0 in $A[i+1, n]$, i.e., the right half of the array.
if $A[i] > 0$, then we will search for 0 in $A[0, i-1]$, i.e., the left half of the array.
if $A[i] = 0$, the algorithm stops.

thus, after each iteration, we are splitting array into half elements. Essentially, we are considering $\frac{n_i}{2}$ elements for evaluation, where n_i is the elements after i^{th} iteration.

hence, eventually the number of steps required for the algorithm to finish will be $O(\log(n))$

As we are constantly reducing our search space and converging in the direction of the keyword (i.e. 0), hence this algorithm will eventually find the solution, provided it exists.

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.
   False

6.2 Discrete and Continuous Distributions

Match the distribution name to its probability density / mass function. Below, $|\boldsymbol{x}| = k$.

(a) Laplace h

(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

   (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) $\mathbb{E}[3X]$
   $\mathbb{E}[3X] = \sum 3xp(x) = 3\sum xp(x) = 3\mathbb{E}[X]$

   $$\implies \mathbb{E}[3X] = 3$$

   (b) $\text{Var}(3X)$
   $Var(3X) = \mathbb{E}[(3X)^2] - (\mathbb{E}[3X])^2$
   $\implies Var(3X) = \mathbb{E}[9X^2] - (3\mathbb{E}[X])^2$
   $\implies Var(3X) = 9\mathbb{E}[X^2] - 9(\mathbb{E}[X])^2$
   $\implies Var(3X) = 9(\mathbb{E}[X^2] - (\mathbb{E}[X])^2)$
   $\implies Var(3X) = 9(Var(X))$

   $$\implies Var(3X) = 9$$

   (c) $\text{Var}(X+3)$
   $\text{Var}(X+3) = \mathbb{E}[(X+3)^2] - (\mathbb{E}[X+3])^2$
   $\implies \text{Var}(X+3) = \mathbb{E}[(X^2+6X+9)] - (\mathbb{E}[X]+\mathbb{E}[3])^2$
   $\implies \text{Var}(X+3) = \mathbb{E}[X^2] + \mathbb{E}[6X] + \mathbb{E}[9] - (\mathbb{E}[X]^2 + 2\mathbb{E}[X]\mathbb{E}[3] + \mathbb{E}[3]^2)$
   $\implies \text{Var}(X+3) = \mathbb{E}[X^2] + \mathbb{E}[X]\mathbb{E}[6] + \mathbb{E}[9] - \mathbb{E}[X]^2 - \mathbb{E}[X]\mathbb{E}[6] - \mathbb{E}[9]$
   $\implies \text{Var}(X+3) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
   $\implies \text{Var}(X+3) = \text{Var}(X)$

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

6.4 Mutual and Conditional Independence

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

   $\mathbb{E}[XY] = \int_x \int_y xyp(x,y)dxdy$
   $\implies \quad = \int_x \int_y xyp(x)p(y)dxdy$, given X and Y are independent
   $\implies \quad = \int_x xp(x)dx \int_y yp(y)dy$
   $\implies \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

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

   $\text{Var}(X+Y) = \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2$
   $\implies \quad = \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2$
   $\implies \quad = \mathbb{E}[X^2] + \mathbb{E}[2XY] + \mathbb{E}[Y^2] - ((\mathbb{E}[X])^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + (\mathbb{E}[Y])^2)$

   given, X and Y are independenct, hence $\mathbb{E}[2XY] = 2\mathbb{E}[X]\mathbb{E}[Y]$
   $\implies \text{Var}(X+Y) = \mathbb{E}[X^2] + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y^2] - (\mathbb{E}[X])^2 - 2\mathbb{E}[X]\mathbb{E}[Y] - (\mathbb{E}[Y])^2$
   $\implies \text{Var}(X+Y) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 + \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$
   $\implies \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$

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. The result of first die will NOT tell us anything about the result of 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?

   No. As in this case, the second die needs to be $\{1, 3, 5\}$.

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.)

   Using law of large numbers,
   $\mathbb{E}[Two\ heads\ occurrences\ after\ 40,000\ repetitions] = \mathbb{E}[\sum_{i=1}^{40,000} I(x_i = H, y_i = H)]$,
   where x_i and y_i are i^{th} occurrences on first and second coins after i^{th} repetition.
   $\implies \quad = (40,000)P(x_i = H, y_i = H)$
   $\implies \quad = (40,000)(\frac{1}{2})(\frac{1}{2})$
   $\implies \quad = 10,000$

   for any i, $P(x_1 = H, x_2 = H) = \frac{1}{4}$, and $P(x_1 \neq H, x_2 \neq H) = \frac{3}{4}$
   $\implies \text{Var}[\sum_{i=1}^{40,000} I(x_i = H, y_i = H)] = \text{Var}[\sum_{i=1}^{40,000} (1 - \frac{1}{4})^2 \frac{1}{4} + (0 - \frac{1}{4})^2 (\frac{3}{4})]$
   $\implies \quad = \text{Var}[\sum_{i=1}^{40,000} (\frac{3}{4})^2 \frac{1}{4} + (\frac{1}{4})^2 (\frac{3}{4})]$
   $\implies \quad = \text{Var}[\sum_{i=1}^{40,000} (\frac{3}{16})]$
   $\implies \quad = 40,000(\frac{3}{16})$
   $\implies \quad = 7500$

   now, standard deviation = $\sqrt{variance}$
   $\implies standarddeviation = \sqrt{7500} \approx 86$

   thus, the expected value for making the required observation is 10,000 with a deviation of $\approx$ 86.

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

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

given, X_i is a Gaussian distribution
$\implies$ $\bar{X}$ is also Gaussian, being sum of Gaussian distribution

$\mathbb{E}[\sqrt{n}\bar{X}] = \mathbb{E}[\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} X_i]$
$\implies \quad = \frac{1}{\sqrt{n}}\mathbb{E}[\sum_{i=1}^{n} X_i]$
$\implies \quad = \frac{1}{\sqrt{n}}\mathbb{E}[X_1 + X_2 + X_3 + ... + X_n]$
$\implies \quad = \frac{1}{\sqrt{n}}(0)$
$\implies \mathbb{E}[\sqrt{n}\bar{X}] = 0$

$\mathbb{V}[\sqrt{n}\bar{X}] = \mathbb{V}[\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} X_i]$
$\implies \quad = \mathbb{V}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i]$
$\implies \quad = (\frac{1}{\sqrt{n}})^2\mathbb{V}[X_1 + X_2 + X_3 + ... + X_n]$
$\implies \quad = \frac{1}{n}\, 1 + 1....1_{n-times}$
$\implies \mathbb{V}[\sqrt{n}\bar{X}] = 1$

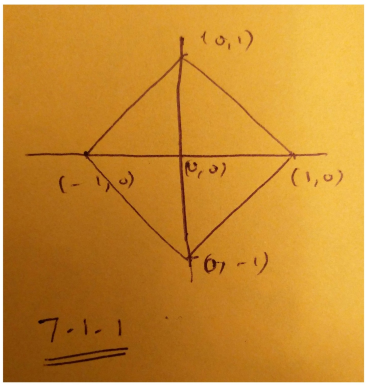
thus, $\sqrt{n}\bar{X} \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1)$

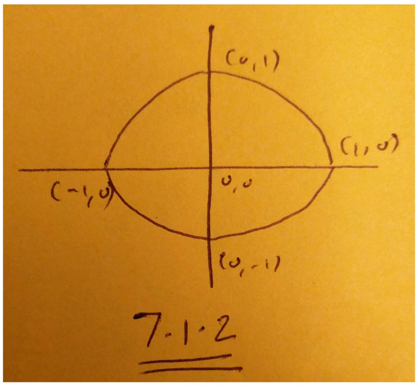
7 Linear algebra [5 pts]

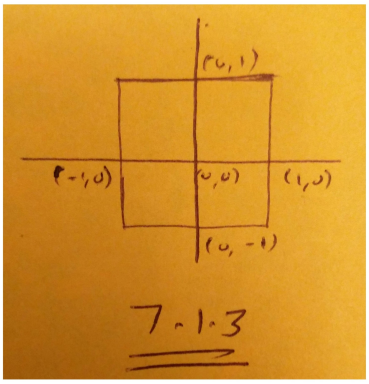
7.1 Norm-enclature

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

1. $||\mathbf{x}||_1 \le 1$ (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
given, the hyperplane is defined by $\mathbf{w}^T\mathbf{x} + b = 0$

$\implies$ the vector perpendicular to the hyperplane would be $\mathbf{w}$
now, assume a point X in the hyperplane.
then, the shortest distance from origin to hyperplane would be the projection of vector $(\mathbf{X} - \mathbf{0})$ on vector $\mathbf{w}$

$\implies d = proj_{\mathbf{w}}\mathbf{X}$
$\implies \quad = ||\frac{\mathbf{X}.\mathbf{w}}{||\mathbf{w}||^2}\mathbf{w}||$

given, $\mathbf{X}$ lies in hyperplane,
$\implies \mathbf{w}^T\mathbf{X} + b = 0$
$\implies \mathbf{X}\mathbf{w} + b = 0$
$\implies \mathbf{X}\mathbf{w} = -b$

$\implies d = ||\frac{-b}{||\mathbf{w}||^2}\mathbf{w}||$
$\implies \quad = \frac{|b|}{||\mathbf{w}||^2}||\mathbf{w}||$
$\implies d = \frac{|b|}{||\mathbf{w}||^2}||\mathbf{w}||$
$\implies d = \frac{|b|}{||\mathbf{w}||_2}$

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
Using analogy from the part 1, let us take two points $X1$ and $X2$ on plane 1 and plane 2 respectively.
now, the closest distance from X_1 to plane 2 will be the projection of $\mathbf{X_2} - \mathbf{X_1}$ on the normal vector of any of the planes (given both the planes are parallel).

$\implies d = proj_{\mathbf{w}}\mathbf{X_2} - \mathbf{X_1}$
$\implies \quad = ||\frac{(\mathbf{X_1}-\mathbf{X_2}).\mathbf{w}}{||\mathbf{w}||^2}\mathbf{w}||$

given, $\mathbf{X_1}$ lies in $\mathbf{w}^T\mathbf{x} + b_1 = 0$ and $\mathbf{X2}$ lies in $\mathbf{w}^T\mathbf{x} + b_2 = 0$,
$\implies \mathbf{w}^T\mathbf{X_1} + b_1 = 0$
$\implies \mathbf{X_1}\mathbf{w} + b_1 = 0$
$\implies \mathbf{X_1}\mathbf{w} = -b_1$

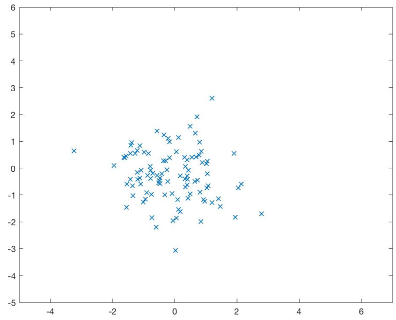
similarly, $\mathbf{X_2}\mathbf{w} = -b_2$

$\implies d = ||\frac{\mathbf{X_1}.\mathbf{w}-\mathbf{X_2}.\mathbf{w}}{||\mathbf{w}||^2}\mathbf{w}||$
$\implies d = ||\frac{-b_1+b_2}{||\mathbf{w}||^2}\mathbf{w}||$
$\implies \quad = \frac{|b_1-b_2|}{||\mathbf{w}||^2}||\mathbf{w}||$
$\implies d = \frac{|b_1-b_2|}{||\mathbf{w}||_2}$

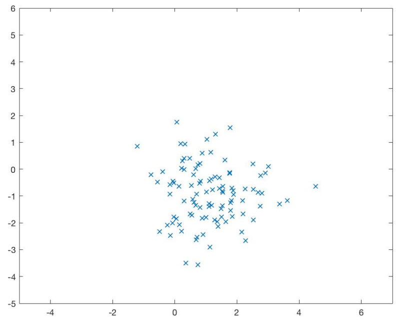
8 Programming Skills - Matlab [5pts]

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.)

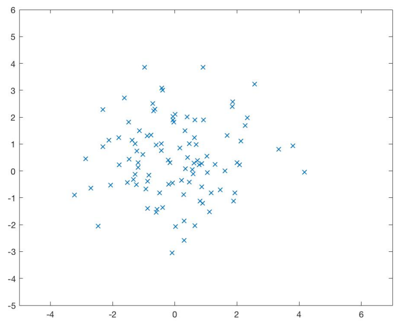
1. Draw 100 samples $\mathbf{x} = [x_1, x_2]^T$ from a 2-dimensional Gaussian distribution with mean $(0, 0)^T$ and identity covariance matrix, i.e., $p(\mathbf{x}) = \frac{1}{2\pi}\exp\left(-\frac{||\mathbf{x}||^2}{2}\right)$, and make a scatter plot (x_1 vs. x_2). For each question below, make each change separately to this distribution.



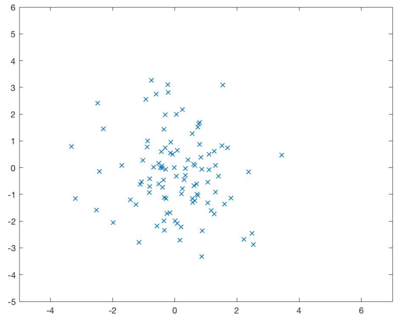
2. Make a scatter plot with a changed mean of $(1, -1)^T$.
   Solution figure goes here.



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 $\begin{pmatrix} 2 & 0.2 \\ 0.2 & 2 \end{pmatrix}$.
   Solution figure goes here.



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