

# HOMWORK 1: BACKGROUND TEST

Name: Deepan Das  
ID: ddas27

## Minimum Background Test [80 pts]

### 1 Vectors and Matrices [20 pts]

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

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

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

$$9*7 + 6*8 = 111$$

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

$$\begin{pmatrix} 9*9 + 8*8 \\ 7*9 + 6*8 \end{pmatrix} = \begin{pmatrix} 145 \\ 111 \end{pmatrix}$$

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

Yes,  $X$  is invertible. Inverse is:

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

4. What is the rank of  $X$ ?

Matrix Rank is 2

### 2 Calculus [20 pts]

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

$$12x^2 - 2x$$

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

$$6z \tan(z)x^{6z-1} - \frac{x^4 \left( \frac{7}{x^4} - \frac{4(7x+z)}{x^5} \right)}{7x+z}$$

### 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?

$$0.5714$$

2. What is the sample variance for this data?

$$0.24489$$

3. What is the probability of observing this data, assuming it was generated by flipping a biased coin with  $p(x = 1) = 0.7, p(x = 0) = 0.3$ .  
 $(0.3)^3 * (0.7)^4 = 0.0064827$

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.

We see that this is essentially a Binomial distribution of a sum of i.i.d. Bernoulli random variables concerning each coin toss. Out of 7 total coin tosses, we have 4 Tails, and so forth. Therefore, to maximize the probability of this random variable which is a Binomial, where  $n = 7$  and  $k = 4$  We try to find a value of  $p(P(x = 1))$  that maximizes the value of  $p^4(1 - p)^3$ .

We see that the value is  $p = 0.5714$

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  $P(A = 0, B = 0)$ ?  
 0.1
- (b) What is  $P(A = 1)$ ?  
 0.5
- (c) What is  $P(A = 0|B = 1)$ ?  
 0.5714
- (d) What is  $P(A = 0 \vee B = 0)$ ?  
 0.15

## 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.

- $f(n) = \frac{n}{2}, g(n) = \log_2(n)$ .  
 $g(n) = O(f(n))$  holds true, not both.
- $f(n) = \ln(n), g(n) = \log_2(n)$ .  
 Both hold true
- $f(n) = n^{100}, g(n) = 100^n$ .  
 $f(n) = O(g(n))$  holds true, not both.

# 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.

Since, we are provided with a sorted array (assume sorted in ascending order), we can use an approach where we keep reducing our search space in the array in successive steps. The best way to do so is to reduce the search space in half at each step. For doing this, we compare the middle element of the array with 0. If the middle element  $mid$  is greater than 0, we know that the element 0 will have to be in the left half of the array. Else, we check the right half of the array. We can keep dividing the resultant array at each step to reduce our search space, until there's just one element left. If the element is not the target (0 in this case), we can safely say that the array doesn't contain the target (0). This is called the binary search approach. Since, we divide the array into 2 parts at each step, there's a maximum of  $\log_2(n)$  comparisons. This, the algorithm runs in  $O(\log(n))$  time. The Algorithm can be summarized as:

- $left = 0, right = n$
- while  $left \leq right$ , check  $mid = (left + right)/2$  is equal to target
- if  $mid == target$ , element found

## 6 Probability and Random Variables [5 pts]

### 6.1 Probability

State true or false. Here  $\Omega$  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  $P(B \cup C) > 0$ ,  $\frac{P(A \cup B \cup C)}{P(B \cup C)} \geq P(A|B \cup C)P(B \cup C)$ .  
False
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,  $|x| = k$ .

- (f)  $f(\mathbf{x}; \Sigma, \mu) = \frac{1}{\sqrt{(2\pi)^k \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right)$
- (g)  $f(x; n, \alpha) = \binom{n}{x} \alpha^x (1 - \alpha)^{n-x}$  for  $x \in \{0, \dots, n\}$ ; 0 otherwise
- (a) Laplace (h). (h)  $f(x; b, \mu) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right)$
- (b) Multinomial (i). (i)  $f(\mathbf{x}; n, \alpha) = \frac{n!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k \alpha_i^{x_i}$  for  $x_i \in \{0, \dots, n\}$  and  $\sum_{i=1}^k x_i = n$ ; 0 otherwise
- (c) Poisson (l). (j)  $f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$  for  $x \in (0, +\infty)$ ; 0 otherwise
- (d) Dirichlet (k). (k)  $f(\mathbf{x}; \alpha) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k x_i^{\alpha_i-1}$  for  $x_i \in (0, 1)$  and  $\sum_{i=1}^k x_i = 1$ ; 0 otherwise
- (e) Gamma (j). (l)  $f(x; \lambda) = \lambda^x \frac{e^{-\lambda}}{x!}$  for all  $x \in \mathbb{Z}^+$ ; 0 otherwise

### 6.3 Mean and Variance

- Consider a random variable which follows a Binomial distribution:  $X \sim \text{Binomial}(n, p)$ .
  - What is the mean of the random variable?  
 $np$
  - What is the variance of the random variable?  
 $np(1 - p)$
- Let  $X$  be a random variable and  $\mathbb{E}[X] = 1$ ,  $\text{Var}(X) = 1$ . Compute the following values:
  - $\mathbb{E}[3X]$   
3
  - $\text{Var}(3X)$   
9
  - $\text{Var}(X + 3)$   
1

### 6.4 Mutual and Conditional Independence

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

$$\mathbb{E}[XY] = \sum_x \sum_y xyp(x, y) = \sum_x \sum_y xyp(x)p(y) = (\sum_x x)(\sum_y y) = \mathbb{E}[X]\mathbb{E}[Y]$$
- If  $X$  and  $Y$  are independent random variables, show that  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ .  
 Hint:  $\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y)$   
 As  $\text{Cov}(X, Y) = 0$ , when  $X$  and  $Y$  are independent variables, we get:  $\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y) = \text{Var}(X) + \text{Var}(Y)$
- 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?  
 Result of second die now becomes dependent on the first die.

### 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.)

According to the Law of Large Numbers, the average of the results obtained from a large number of trials is close to the expected value. In this case, the probability of two heads follows a Binomial Distribution like:

$$\mathbb{P}(\sum_{i=1}^n Y_i = k) = {}^n C_k p^k (1-p)^{(n-k)}$$

where  $k = 2$  in our case, as we want two heads. Put simply,  $\mathbb{P}(HH) = \frac{1}{4}$ , as there are 4 possible outcomes  $= HH, HT, TH, TT$ .

Considering this, at number of repetitions = 40,000 and by following the law of large numbers, the number of times we have two heads ( $HH$ ) as the output is:

$$\text{Number of HH} = \mathbb{P}(HH) * 40,000 = 10,000$$

Also, from the law of large numbers, the standard deviation of the random variable indicating two heads successively is given by:

$$\text{Std.Deviation} = \sqrt{\frac{\text{Var}(HH)}{n}} = \frac{1/2}{\sqrt{40,000}} = 0.0000125. \text{ Therefore, considering two standard deviations, the number of times we see } HH \text{ is in the interval } [9999.999975, 10000.0000125]$$

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

$$\sqrt{n}\bar{X} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1)$$

From the Central Limit Theorem, we know that when Independent random variables are added, their normalized sum tends towards a normal distribution. Moreover, by Law of Large Numbers, for infinitesimally large  $n$ , we have,

$$\mathbb{E}(\sqrt{n}\bar{X}) = \frac{\sqrt{n}}{n} \sum_{i=1}^n \mathbb{E}(X_i) = 0, \text{ because } X_i \sim \mathcal{N}(0, 1). \text{ Also, in lieu of the same distribution property,}$$

$$\text{Var}(\sqrt{n}\bar{X}) = n * \text{Var}(\frac{1}{n} \sum_{i=1}^n X_i) = \frac{1}{n} * \text{Var}(\sum_{i=1}^n X_i) = \frac{n}{n} = 1.$$

Thus,

$$\sqrt{n}\bar{X} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1)$$

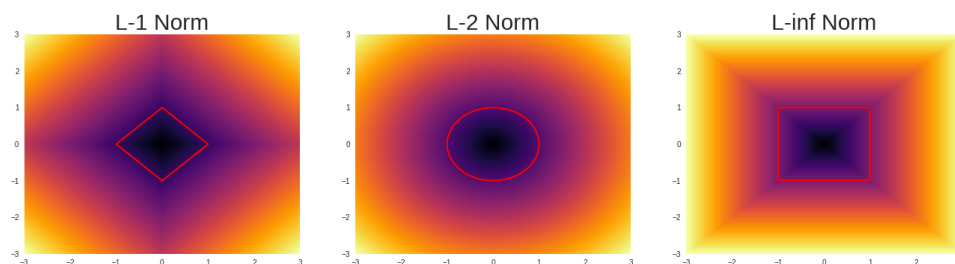
## 7 Linear algebra [5 pts]

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

[Solution figure goes here.](#) Boundary shown in red. Contours provided for greater understanding.



## 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  $\frac{|b|}{\|\mathbf{w}\|_2}$ .  
Suppose we find a point  $X$  on the hyperplane. We essentially need to find the projection of the vector from  $X$  to  $X_0(0, 0)$  on the perpendicular to the hyperplane  $\mathbf{w}^T \mathbf{x} + b = 0$ . We know that the perpendicular to the hyperplane  $\mathbf{w}^T \mathbf{x} + b = 0$  is  $\mathbf{w}$ . Thus,

$d = \text{projection of } X - X_0 \text{ on } \mathbf{w}$

Let  $(X - X_0) = a$

$$d = \|a\|_2 \cos \theta = \|a\|_2 \frac{a \cdot \mathbf{w}}{\|a\|_2 \|\mathbf{w}\|_2} = \frac{a \cdot \mathbf{w}}{\|\mathbf{w}\|_2}$$

Substituting  $a = (X - X_0)$

$$d = \frac{\mathbf{w} \cdot (X - X_0)}{\|\mathbf{w}\|_2} = \frac{\mathbf{w}^T X - \mathbf{w}^T X_0}{\|\mathbf{w}\|_2} = \frac{|b|}{\|\mathbf{w}\|_2}$$

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).

We can find the distances  $d_1$  and  $d_2$  for both the hyperplanes respectively and notice that our final solution  $d = d_1 - d_2$ .

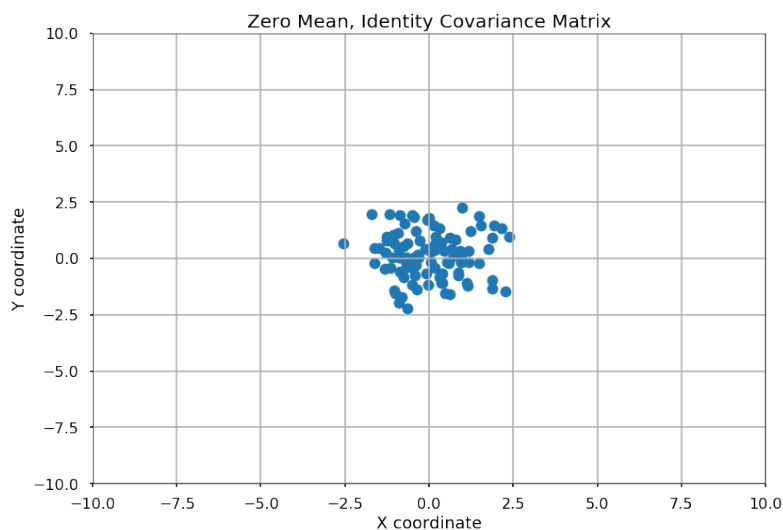
$$d_1 = \frac{|b_1|}{\|\mathbf{w}\|_2} \quad d_2 = \frac{|b_2|}{\|\mathbf{w}\|_2} \quad d = d_1 - d_2 = \frac{|b_1 - b_2|}{\|\mathbf{w}\|_2}$$

## 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.)

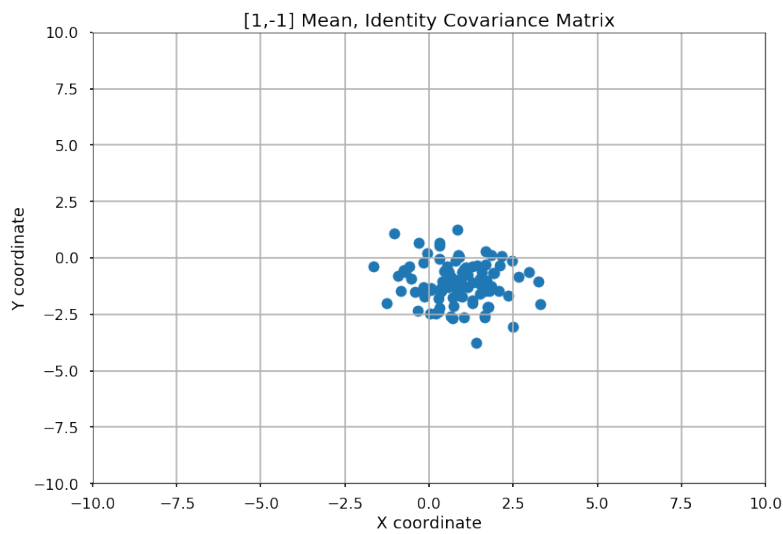
1. Draw 100 samples  $\mathbf{x} = [x_1, x_2]^T$  from a 2-dimensional Gaussian distribution with mean  $(0, 0)^T$  and identity covariance matrix, i.e.,  $p(\mathbf{x}) = \frac{1}{2\pi} \exp\left(-\frac{\|\mathbf{x}\|^2}{2}\right)$ , and make a scatter plot ( $x_1$  vs.  $x_2$ ). For each question below,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

Points centred around [0,0].



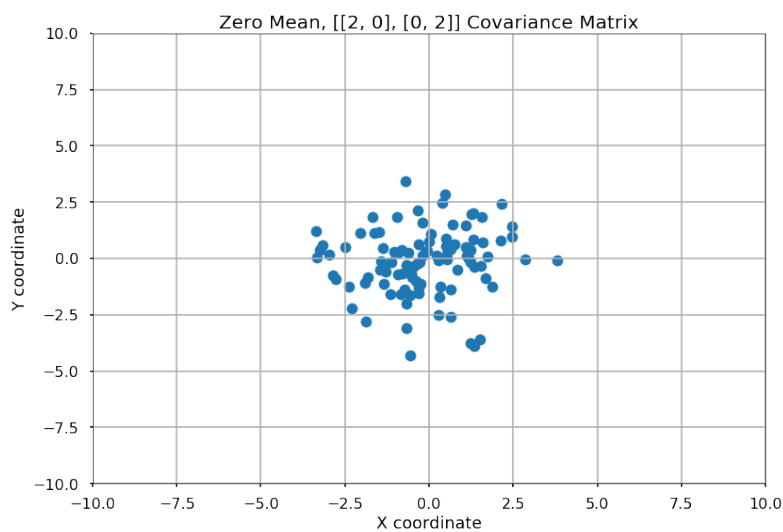
2. Make a scatter plot with a changed mean of  $(1, -1)^T$ .

Same as before, but points are now centered around  $[1, -1]$



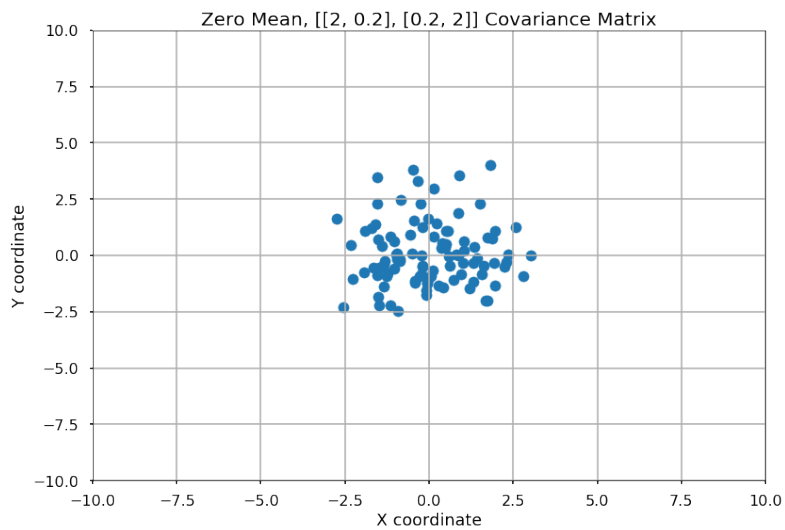
3. Make a scatter plot with a changed covariance matrix of  $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ .

Same as 1, but doubled spread or variance across both dimensions.



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

Same as 4, but there's a small positive correlation with  $x_1$  and  $x_2$ .



5. Make a scatter plot with a changed covariance matrix of  $\begin{pmatrix} 2 & -0.2 \\ -0.2 & 2 \end{pmatrix}$ .  
Same as 4, but there's a small negative correlation with  $x_1$  and  $x_2$ .

