

# HOMework 1: BACKGROUND TEST

Wenyan Zhou  
9077005776

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

$$\mathbf{y}^T \mathbf{z} = (9 \ 8) \begin{pmatrix} 7 \\ 6 \end{pmatrix} = 111$$

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

$$X\mathbf{y} = \begin{pmatrix} 9 & 8 \\ 7 & 6 \end{pmatrix} \begin{pmatrix} 9 \\ 8 \end{pmatrix} = \begin{pmatrix} 145 \\ 111 \end{pmatrix}$$

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

Yes.

Let

$$X^{-1}X = I,$$

then

$$X^{-1} = \begin{pmatrix} -3 & 4 \\ 3.5 & -4.5 \end{pmatrix}$$

4. What is the rank of  $X$ ?

$$|X| = \begin{vmatrix} 9 & 8 \\ 7 & 6 \end{vmatrix} = -2 \neq 0$$

Therefore, 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  $x$ ?

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

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

$$\begin{aligned}\frac{\partial y}{\partial x} &= 6z \tan(z) x^{6z-1} - \frac{x^4}{7x+z} \frac{7x^4 - 4x^3(7x+z)}{x^8} \\ &= 6z \tan(z) x^{6z-1} - \frac{-21x^4 - 4zx^3}{7x^5 + zx^4} \\ &= 6z \tan(z) x^{6z-1} + \frac{21x + 4z}{7x^2 + zx}\end{aligned}$$

### 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?

$$\bar{x} = (0 + 1 + 1 + 0 + 0 + 1 + 1)/7 = \frac{4}{7}$$

2. What is the sample variance for this data?

$$S^2 = [3 \times (0 - \frac{4}{7})^2 + 4 \times (1 - \frac{4}{7})^2]/6 = \frac{2}{7}$$

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

$$p(3H, 4T) = \binom{7}{3} \times 0.3^3 \times 0.7^4 = 0.2268945$$

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.

Suppose  $p(x=1) = p$ ,

the probability of observing  $S$  is  $p(3H, 4T) = \binom{7}{3} \times p^4(1-p)^3$ .

To maximize the value ( $p \neq 0$ ),  $p = \frac{4}{7}$ .

Then,  $p(3H, 4T) = 0.2937552$

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  $P(A=1)$ ?

$$P(A=1) = P(A=1, B=0) + P(A=1, B=1) = 0.2 + 0.3 = 0.5$$

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

$$P(A = 0|B = 1) = \frac{P(A = 0, B = 1)}{P(B = 1)} = 0.4/(0.4 + 0.3) = \frac{4}{7}$$

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

$$P(A = 0 \vee B = 0) = 1 - P(A = 1, B = 1) = 1 - 0.3 = 0.7$$

## 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.  $f(n) = \frac{n}{2}$ ,  $g(n) = \log_2(n)$ .  
 $g(n) = O(f(n))$

Because

$$\lim_{n \rightarrow \infty} \sup \left| \frac{g(n)}{f(n)} \right| = \lim_{n \rightarrow \infty} \sup \left| \frac{\log_2(n)}{n/2} \right| = \lim_{n \rightarrow \infty} \sup \left| \frac{(1/n)(1/\ln 2)}{1/2} \right| = \lim_{n \rightarrow \infty} \sup \left| \frac{2}{n \ln 2} \right| < \infty$$

$$\lim_{n \rightarrow \infty} \sup \left| \frac{f(n)}{g(n)} \right| = \lim_{n \rightarrow \infty} \sup \left| \frac{n \ln 2}{2} \right| = \infty$$

2.  $f(n) = \ln(n)$ ,  $g(n) = \log_2(n)$ .  
 Both  $f(n) = O(g(n))$  and  $g(n) = O(f(n))$   
 Because

$$\lim_{n \rightarrow \infty} \sup \left| \frac{f(n)}{g(n)} \right| = \lim_{n \rightarrow \infty} \sup \left| \frac{\ln n}{\log_2(n)} \right| = \lim_{n \rightarrow \infty} \sup \left| \frac{1/n}{(1/n)(1/\ln 2)} \right| = \ln 2 < \infty$$

$$\lim_{n \rightarrow \infty} \sup \left| \frac{g(n)}{f(n)} \right| = \frac{1}{\ln 2} < \infty$$

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

$$\lim_{n \rightarrow \infty} \sup \left| \frac{f(n)}{g(n)} \right| = \lim_{n \rightarrow \infty} \sup \left| \frac{n^{100}}{100^n} \right| = \lim_{n \rightarrow \infty} \sup \left| \frac{100!}{(\ln 100)^{100} 100^n} \right| < \infty$$

$$\lim_{n \rightarrow \infty} \sup \left| \frac{g(n)}{f(n)} \right| = \lim_{n \rightarrow \infty} \sup \left| \frac{100^n}{n^{100}} \right| = \lim_{n \rightarrow \infty} \sup \left| \frac{(\ln 100)^{100} 100^n}{100!} \right| = \infty$$

# 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.

Begin with dividing the whole array in half. If the middle of the interval is 0, we find it; else if 0 is less than the number in the middle of the interval, narrow the interval to the lower half; otherwise, narrow it to the upper half. Repeat the above steps until 0 is found. To justify the running time  $t$ ,  $2^t = n$ , and  $t = \log_2(n) = \log(n)/\log(2) = O(\log(n))$

## 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.

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

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

False.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

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.

$$\frac{P(A \cup B \cup C)}{P(B \cup C)} \geq P(A \cup B \cup C) \geq P(A \cap (B \cup C)) = P(A|B \cup C)P(B \cup C)$$

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.

Suppose  $B \subseteq A$ , and  $P(B) < P(A)$ , then  $P(B|A^c) + P(B|A) = P(B|A) = P(AB)/P(A) < 1$

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.

$P(\bigcap_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$ , which indicates  $P(A_i) = 0$  for all  $A_i$ . then  $\{A_i\}_{i=1}^n$  are mutually independent.

### 6.2 Discrete and Continuous Distributions

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

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

$$\mathbb{E}[X] = np$$

(b) What is the variance of the random variable?

$$\text{Var}(X) = np(1 - p)$$

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

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

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

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

$$\text{Var}(3X) = 9\text{Var}(X) = 9$$

(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  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ .

Take continuous variables as an example, it is the same for discrete variables.

$$\begin{aligned} \mathbb{E}[XY] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x)f(y) dx dy \\ &= \int_{-\infty}^{+\infty} xf(x) dx \int_{-\infty}^{+\infty} yf(y) dy = \mathbb{E}[X]\mathbb{E}[Y] \end{aligned}$$

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

$$\begin{aligned}\text{Var}(X + Y) &= \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y) \\ &= \text{Var}(X) + 2[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)] + \text{Var}(Y) \\ &= \text{Var}(X) + \text{Var}(Y)\end{aligned}$$

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. Because the two events 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. The third event indicates that the second die's result is 1 given the first die's result is 1, so that the sum of the two results is even.

## 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 LLN,  $\bar{X}_n \xrightarrow{P} \mu$  when  $n \rightarrow \infty$ , because  $\mu = \frac{1}{4}$ , the number of two heads results is close to 10,000 for 40,000 repetitions.

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

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

According to CLT,  $\sqrt{n}(\bar{X} - \mu) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma^2)$ ,  $\mathbb{E}(X_i) = \mu = 0$ ,  $\text{Var}(X_i) = \sigma^2 = 1$ , thus proved.

## 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:

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

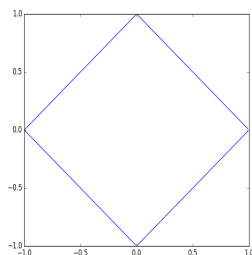


fig 7.1.1

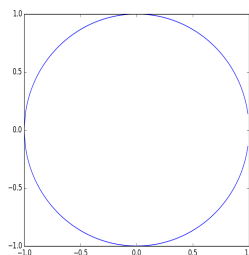


fig 7.1.2

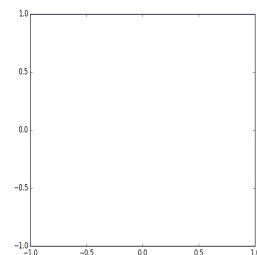


fig 7.1.3

## 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  $\frac{|b|}{\|\mathbf{w}\|_2}$ .  
True.

Suppose the projection of origin  $\mathbf{x}_0 = (0, 0, \dots, 0)_n$  on the hyperplane is  $\mathbf{x}_1 = (x_1^1, x_1^2, \dots, x_1^n)$ , then  $\mathbf{x}_1$  satisfies  $\mathbf{w}^T \mathbf{x}_1 + b = 0$ , the normal vector of the hyperplane is  $\mathbf{w}$ , the required smallest Euclidean distance is  $d$ .

$$\begin{aligned} |\mathbf{w}^T \overrightarrow{\mathbf{x}_0 \mathbf{x}_1}| &= |\mathbf{w}^T| |\overrightarrow{\mathbf{x}_0 \mathbf{x}_1}| = \|\mathbf{w}\|_2 d \\ \mathbf{w}^T \overrightarrow{\mathbf{x}_0 \mathbf{x}_1} &= w_1 x_1^1 + w_2 x_1^2 + \dots w_n x_1^n \end{aligned}$$

Therefore,

$$\begin{aligned} \|\mathbf{w}\|_2 d &= |w_1 x_1^1 + w_2 x_1^2 + \dots w_n x_1^n| = |\mathbf{w}^T \mathbf{x}_1| = |b| \\ d &= \frac{|b|}{\|\mathbf{w}\|_2} \end{aligned}$$

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.

Suppose  $\mathbf{x}_1 = (x_1^1, x_1^2, \dots, x_1^n)$  is an arbitrary point on hyperplane  $\mathbf{w}^T \mathbf{x} + b_1 = 0$  satisfying  $\mathbf{w}^T \mathbf{x}_1 + b_1 = 0$ . The projection of  $\mathbf{x}_1$  on the hyperplane  $\mathbf{w}^T \mathbf{x} + b_2 = 0$  is  $\mathbf{x}_2 = (x_2^1, x_2^2, \dots, x_2^n)$ , which satisfies  $\mathbf{w}^T \mathbf{x}_2 + b_2 = 0$ , the normal vector of the hyperplane is  $\mathbf{w}$ , the required smallest Euclidean distance is  $d$ .

$$\begin{aligned} |\mathbf{w}^T \overrightarrow{\mathbf{x}_1 \mathbf{x}_2}| &= |\mathbf{w}^T| |\overrightarrow{\mathbf{x}_1 \mathbf{x}_2}| = \|\mathbf{w}\|_2 d \\ \mathbf{w}^T \overrightarrow{\mathbf{x}_1 \mathbf{x}_2} &= w_1(x_2^1 - x_1^1) + w_2(x_2^2 - x_1^2) + \dots w_n(x_2^n - x_1^n) \\ &= w_1 x_2^1 + \dots + w_n x_2^n - (w_1 x_1^1 + \dots + w_n x_1^n) \\ &= -b_2 - (-b_1) = b_1 - b_2 \end{aligned}$$

Therefore,

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

## 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.

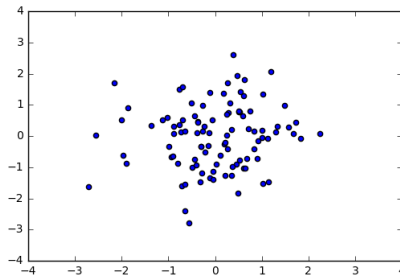


fig 8.1

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

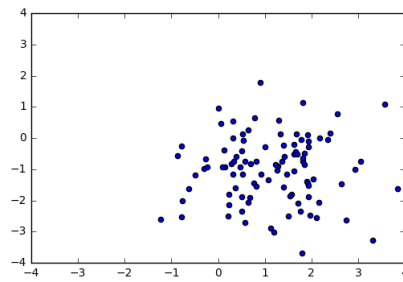


fig 8.2

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

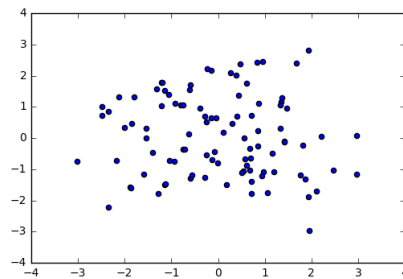


fig 8.3

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

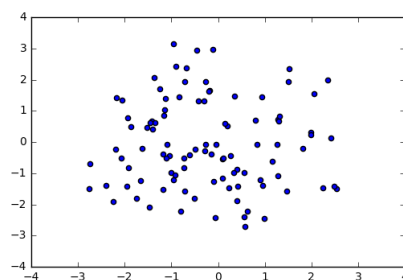


fig 8.4

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



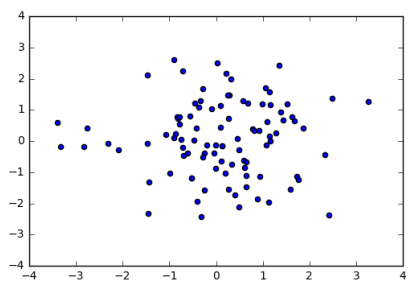


fig 8.5