

# CS M146 Fall 2023 Homework 0

UCLA ID: [205452321]

Name: [Ken DENG]

Collaborators: [list all the people you worked with]

By turning in this assignment, I agree by the UCLA honor code and declare that all of this is my own work.

## Problem 1

Consider  $y = x \sin(z)e^{-x}$ . What is the partial derivative of  $y$  with respect to  $x$ ?

$$\begin{aligned}\frac{\partial y}{\partial x} &= \frac{\partial(\sin(z)xe^{-x})}{\partial x} \\ &= \sin(z)e^{-x} - \sin(z)xe^{-x} \\ &= (1-x)\sin(z)e^{-x}\end{aligned}$$

## Problem 2

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

$$\mathbf{X} = \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad \mathbf{z} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

i. What is the inner product  $\mathbf{y}^T \mathbf{z}$ ?

$$\begin{aligned}\mathbf{y}^T \mathbf{z} &= (1 \ 3) \begin{pmatrix} 2 \\ 3 \end{pmatrix} \\ &= 1 \cdot 2 + 3 \cdot 3 \\ &= 2 + 9 \\ &= 11\end{aligned}$$

ii. What is the product  $\mathbf{Xy}$ ?

$$\begin{aligned}\mathbf{Xy} &= \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} \\ &= 2 \cdot 1 + 4 \cdot 3 + 1 \cdot 1 + 3 \cdot 3 \\ &= 2 + 12 + 1 + 9 \\ &= 24\end{aligned}$$

iii. Is  $\mathbf{X}$  invertible? If so, give the inverse; if not, explain why not.

Yes,  $\mathbf{X}$  is invertible.

$$\begin{pmatrix} 2 & 4 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & \frac{1}{2} & 0 \\ 1 & 3 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & \frac{1}{2} & 0 \\ 0 & 1 & -\frac{1}{2} & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & \frac{3}{2} & -2 \\ 0 & 1 & -\frac{1}{2} & 1 \end{pmatrix}$$

Therefore,

$$\mathbf{X}^{-1} = \begin{pmatrix} \frac{3}{2} & -2 \\ -\frac{1}{2} & 1 \end{pmatrix}$$

iv. What is the rank of  $\mathbf{X}$ ?

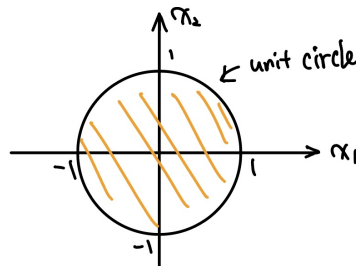
$$\begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

Therefore,

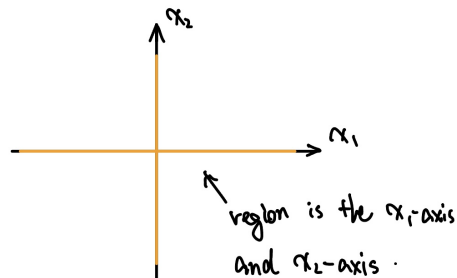
$$\text{rank}(\mathbf{X}) = 2$$

(b) Draw the regions corresponding to vectors  $\mathbf{x} \in \mathbb{R}^2$  with following norms (you can hand draw or use software for this question):

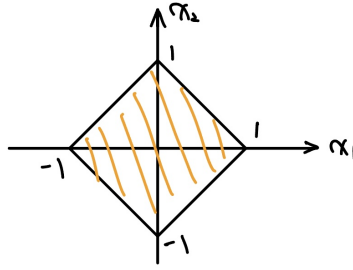
i.  $\|\mathbf{x}\|_2 \leq 1$  (Recall  $\|\mathbf{x}\|_2 = \sqrt{\sum_i x_i^2}$ .)



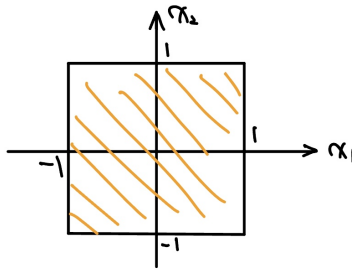
ii.  $\|\mathbf{x}\|_0 \leq 1$  (Recall  $\|\mathbf{x}\|_0 = \sum_{i: x_i \neq 0} 1$ .)



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



iv.  $\|\mathbf{x}\|_\infty \leq 1$  (Recall  $\|\mathbf{x}\|_\infty = \max_i |x_i|$ .)



(c)

i. Give the definition of the eigenvalues and the eigenvectors of a square matrix.

Let  $A$  be an  $n \times n$  matrix.

An **eigenvector** of  $A$  is a non-zero vector  $v$  in  $\mathbb{R}^n$  such that  $Av = \lambda v$ , for some scalar  $\lambda$ .

An **eigenvalue** of  $A$  is a scalar  $\lambda$  such that equation  $Av = \lambda v$  has a nontrivial solution.

If  $Av = \lambda v$  for  $v \neq 0$ ,  $\lambda$  is the **eigenvalue** for  $v$ , and  $v$  is the **eigenvector** for  $\lambda$ .

ii. Find the eigenvalues and eigenvectors of

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{pmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{pmatrix} \\ &= 4 - 4\lambda + \lambda^2 - 1 \\ &= (\lambda - 1)(\lambda - 3) \end{aligned}$$

Hence, we have eigenvalues are  $\lambda_1 = 1, \lambda_2 = 3$ .

For  $\lambda_1 = 1$ :

$$\begin{aligned}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ 2x_1 + x_2 &= x_1 \\ x_1 + 2x_2 &= x_2 \\ x_1 &= -x_2\end{aligned}$$

So we have our eigenvalue for  $\lambda_1 = 1$  as  $v_1 = a \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ , where  $a$  is any non-zero real number.

For  $\lambda_2 = 3$ :

$$\begin{aligned}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3x_1 \\ 3x_2 \end{bmatrix} &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ 2x_1 + x_2 &= 3x_1 \\ x_1 + 2x_2 &= 3x_2 \\ x_1 &= x_2\end{aligned}$$

So we have our eigenvalue for  $\lambda_2 = 3$  as  $v_2 = b \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ , where  $b$  is any non-zero real number.

iii. For any positive integer  $k$ , show that the eigenvalues of  $\mathbf{A}^k$  are  $\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k$ , the  $k^{th}$  powers of the eigenvalues of matrix  $\mathbf{A}$ , and that each eigenvector of  $\mathbf{A}$  is still an eigenvector of  $\mathbf{A}^k$ .

For an arbitrary  $\lambda_i$ , where  $i \in \{1, 2, \dots, n\}$ , we have  $\mathbf{A}v_i = \lambda_i v_i$ .

Then,  $\mathbf{A}\mathbf{A}\lambda_i v_i = \mathbf{A}\lambda_i v_i = \lambda_i \mathbf{A}v_i = \lambda_i \lambda_i v_i$ .

Similarly,  $\mathbf{A}^k v_i = \mathbf{A}^{k-1} \mathbf{A}v_i = \mathbf{A}^{k-1} \lambda_i v_i = \lambda_i \mathbf{A}^{k-1} v_i = \lambda_i^2 \mathbf{A}^{k-2} v_i = \lambda_i^3 \mathbf{A}^{k-3} v_i = \dots = \lambda_i^k v_i$ .

Therefore, according to the definition of eigenvalue and eigenvector, we see that the eigenvalues of  $\mathbf{A}^k$  are  $\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k$ , the  $k^{th}$  powers of the eigenvalues of matrix  $\mathbf{A}$ , and that each eigenvector of  $\mathbf{A}$  is still an eigenvector of  $\mathbf{A}^k$ .

Proved.  $\square$

(d) Consider the vectors  $\mathbf{x}$  and  $\mathbf{a}$  and the symmetric matrix  $\mathbf{A}$ .

i. What is the first derivative of  $\mathbf{a}^T \mathbf{x}$  with respect to  $\mathbf{x}$ ?

We denote  $f(x) = \mathbf{a}^T \mathbf{x} = \sum_{i=1}^n a_i x_i$ .

$$\text{So } \frac{df}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \dots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}.$$

$$\text{Then, } \frac{\partial f}{\partial x_j} = \frac{\partial(\sum_{i=1}^n a_i x_i)}{\partial x_j} = a_j.$$

$$\text{Hence, } \frac{d(\mathbf{a}^T \mathbf{x})}{d\mathbf{x}} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix} = \mathbf{a}^T.$$

ii. What is the first derivative of  $\mathbf{x}^T \mathbf{A} \mathbf{x}$  with respect to  $\mathbf{x}$ ? What is the second derivative?

First derivative:

$$\text{We denote } f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j.$$

$$\text{So } \frac{df}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \dots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}.$$

Then,

$$\begin{aligned} \frac{\partial f}{\partial x_k} &= \frac{\partial(\sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j)}{\partial x_k} = \frac{\partial(\sum_{i=1}^n A_{ik} x_i x_k + \sum_{j=1}^n A_{kj} x_k x_j)}{\partial x_k} \\ &= \sum_{i=1}^n A_{ik} x_i + \sum_{j=1}^n A_{kj} x_j \\ &= 2 \sum_{i=1}^n A_{ki} x_i. \end{aligned}$$

$$\text{Therefore, } \frac{df}{d\mathbf{x}} = \begin{bmatrix} 2 \sum_{i=1}^n A_{1i} x_i \\ 2 \sum_{i=1}^n A_{2i} x_i \\ \dots \\ 2 \sum_{i=1}^n A_{ni} x_i \end{bmatrix} = 2 \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = 2\mathbf{A}\mathbf{x}.$$

Second derivative:

$$\frac{d^2 f}{d\mathbf{x}^2} = \frac{d(2\mathbf{A}\mathbf{x})}{d\mathbf{x}} = 2\mathbf{A}.$$

(e)

i. Show that the vector  $\mathbf{w}$  is orthogonal to the line  $\mathbf{w}^T \mathbf{x} + b = 0$ . (Hint: Consider two points  $\mathbf{x}_1, \mathbf{x}_2$  that lie on the line. What is the inner product  $\mathbf{w}^T(\mathbf{x}_1 - \mathbf{x}_2)$ ?)

Consider 2 points  $\mathbf{x}_1, \mathbf{x}_2$  lying on the line.

Then, the inner product  $\mathbf{w}^T(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{w}^T \mathbf{x}_1 - \mathbf{w}^T \mathbf{x}_2 = -b + b = 0$ .

Since the inner product of  $\mathbf{w}$  and  $\mathbf{x}_1 - \mathbf{x}_2$  is 0, we get that  $\mathbf{w}$  is orthogonal to  $\mathbf{x}_1 - \mathbf{x}_2$ , which means  $\mathbf{w}$  is orthogonal to the line  $\mathbf{w}^T \mathbf{x} + b = 0$ .

Proved .  $\square$ .

ii. Argue that the distance from the origin to the line  $\mathbf{w}^T \mathbf{x} + b = 0$  is  $\frac{b}{\|\mathbf{w}\|_2}$ .

From i, we see that  $\mathbf{w}$  is orthogonal to the line. Hence, if we denote the projection of origin to the line as  $\mathbf{x}_p$ , we see that  $\mathbf{x}_p = m\mathbf{w}$ , for some real number  $m$ .

Then,  $\mathbf{w}^T \mathbf{x}_p + b = 0$ , and thus  $\mathbf{w}^T(m\mathbf{w}) + b = 0$ . So  $m = \frac{b}{\mathbf{w}^T \mathbf{w}}$ , and  $\mathbf{x}_p = \frac{b\mathbf{w}}{\mathbf{w}^T \mathbf{w}} = \frac{b}{\mathbf{w}^T}$ .

Therefore, the distance is  $\|\mathbf{x}_p\|_2 = \|\frac{b}{\mathbf{w}^T}\|_2 = \sqrt{\frac{b^2}{\mathbf{w}^T \mathbf{w}}} = \frac{b}{\|\mathbf{w}\|_2}$ .

Proved .  $\square$ .

## Problem 3

(a) Consider a sample of data  $S$  obtained by flipping a coin five times.  $X_i, i \in \{1, \dots, 5\}$  is a random variable that takes a value 0 when the outcome of coin flip  $i$  turned up heads, and 1 when it turned up tails. Assume that the outcome of each of the flips does not depend on the outcomes of any of the other flips. The sample obtained  $S = (X_1, X_2, X_3, X_4, X_5) = (1, 1, 0, 1, 0)$ .

i. What is the sample mean for this data?

$$\bar{S} = (1 + 1 + 0 + 1 + 0)/5 = 3/5.$$

ii. What is the unbiased sample variance?

$$s^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2 = 1/4 \cdot ((1-0.6)^2 + (1-0.6)^2 + (0-0.6)^2 + (1-0.6)^2 + (0-0.6)^2) = 0.3.$$

iii. What is the probability of observing this data assuming that a coin with an equal probability of heads and tails was used? (i.e., The probability distribution of  $X_i$  is  $P(X_i = 1) = 0.5, P(X_i = 0) = 0.5$ .)

$$0.5^5 = 1/32.$$

iv. Note the probability of this data sample would be greater if the value of the probability of heads  $P(X_i = 1)$  was not 0.5 but some other value. What is the value that maximizes the probability of the sample  $S$ ? [Optional: Can you prove your answer is correct?]

We let  $P(X_i = 1) = p$ . Then we want to maximize the value of  $p^3(1 - p)^2$ .

$$\text{Then } \frac{d(p^3(1 - p)^2)}{dp} = \frac{d(p^5 - 2p^4 + p^3)}{dp} = 5p^4 - 8p^3 + 3p^2 = p^2(5p - 3)(p - 1) = 0.$$

We have  $p_1 = 0, p_2 = 3/5, p_3 = 1$ .

We see when  $p_1 = 0$  or  $p_3 = 1$ , the probability is 0, which is minimized, so we exclude these two points.

Now we consider  $p_2 = 3/5$ .

The second derivative is  $\frac{d(5p^4 - 8p^3 + 3p^2)}{dp} = 20p^3 - 24p^2 + 6p = -0.72 < 0$ . So this point is surely a local maximum.

Thus,  $P(X_i = 1) = 3/5$  maximizes the probability of the sample  $S$ .

v. Given the following joint distribution between  $X$  and  $Y$ , what is  $P(X = T|Y = b)$ ?

$$P(X = T|Y = b) = \frac{P(X = T \cap Y = b)}{P(Y = b)} = \frac{0.1}{0.1 + 0.15} = 2/5.$$

(b) Match the distribution name to its formula.

Gaussian  $\rightarrow \frac{1}{\sqrt{(2\pi)\sigma^2}} \exp(-\frac{1}{2\sigma^2}(x - \mu)^2)$ .

Exponential  $\rightarrow \lambda e^{-\lambda x}$  when  $x \geq 0$ ; 0 otherwise.

Uniform  $\rightarrow \frac{1}{b-a}$  when  $a \leq x \leq b$ ; 0 otherwise.

Bernoulli  $\rightarrow p^x(1 - p)^{1-x}$ , when  $x \in \{0, 1\}$ ; 0 otherwise.

Binomial  $\rightarrow \binom{n}{x} p^x (1 - p)^{n-x}$ .

(c) What is the mean and variance of a *Bernoulli*( $p$ ) random variable?

$$E[\text{Bernoulli}(p)] = 1p + 0p = p.$$

$$\text{Var}(\text{Bernoulli}(p)) = E[X^2] - (E[X])^2 = (1^2p + 0^2p) - p^2 = p - p^2.$$

(d) If the variance of a zero-mean random variable  $X$  is  $\sigma^2$ , what is the variance of  $2X$ ? What about the variance of  $X + 2$ ?

$$\text{Var}(X) = \sigma^2 = E[X^2] - (E[X])^2.$$

$$\text{Var}(2X) = E[(2X)^2] - (E[2X])^2 = E[4X^2] - (2E[X])^2 = 4(E[X^2] - (E[X])^2) = 4\sigma^2.$$

$$\text{Var}(X + 2) = E[(X + 2)^2] - (E[X + 2])^2 = (E[X^2] + E[4X] + E[4]) - (E[X] + E[2])^2 = E[X^2] + 4E[X] + 4 - (E[X])^2 - 4E[X] - 4 = E[X^2] - (E[X])^2 = \sigma^2.$$

## Problem 4

**Big-O notation** For each pair  $(f, g)$  of functions below, list which of the following are true:  $f(n) = O(g(n))$ ,  $g(n) = O(f(n))$ , or both. Briefly justify your answers.

(a)  $f(n) = \ln(n)$ ,  $g(n) = \lg(n)$ . Note that  $\ln$  denotes log to the base  $e$  and  $\lg$  denotes log to the base 2.

Both.

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = \frac{\ln(n)}{\lg(n)} = \frac{\log_e(n)}{\log_2(n)} = \frac{\log_2(n)/\log_2(e)}{\log_2(n)} = \frac{1}{\log_2(e)} < \infty.$$

$$\text{So } f(n) = O(g(n)).$$

$$\lim_{n \rightarrow \infty} \frac{g(n)}{f(n)} = \frac{\lg(n)}{\ln(n)} = \log_2(e) < \infty.$$

$$\text{So } g(n) = O(f(n)).$$

(b)  $f(n) = 3^n$ ,  $g(n) = n^{10}$

$$g(n) = O(f(n)).$$

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = \frac{3^n}{n^{10}}.$$

$$\text{Now we use ratio test: } \lim_{n \rightarrow \infty} \frac{3^{n+1}}{(n+1)^{10}} \cdot \frac{n^{10}}{3^n} = 3 \cdot \lim_{n \rightarrow \infty} \frac{n^{10}}{(n+1)^{10}} = 3 > 1.$$

$$\text{Therefore, } \lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = \infty.$$

$$\text{So } f(n) \neq O(g(n)).$$



$$\lim_{n \rightarrow \infty} \frac{g(n)}{f(n)} = \frac{n^{10}}{3^n}.$$

Now we use ratio test:  $\lim_{n \rightarrow \infty} \frac{(n+1)^{10}}{3^{n+1}} \cdot \frac{3^n}{n^{10}} = \frac{1}{3} \cdot \lim_{n \rightarrow \infty} \frac{(n+1)^{10}}{n^{10}} = \frac{1}{3} < 1$ .

Therefore,  $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} < \infty$ .

So  $g(n) = O(f(n))$ .

(c)  $f(n) = 3^n$ ,  $g(n) = 2^n$

$$g(n) = O(f(n)).$$

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = \lim_{n \rightarrow \infty} \frac{3^n}{2^n} = \lim_{n \rightarrow \infty} \left(\frac{3}{2}\right)^n = \infty.$$

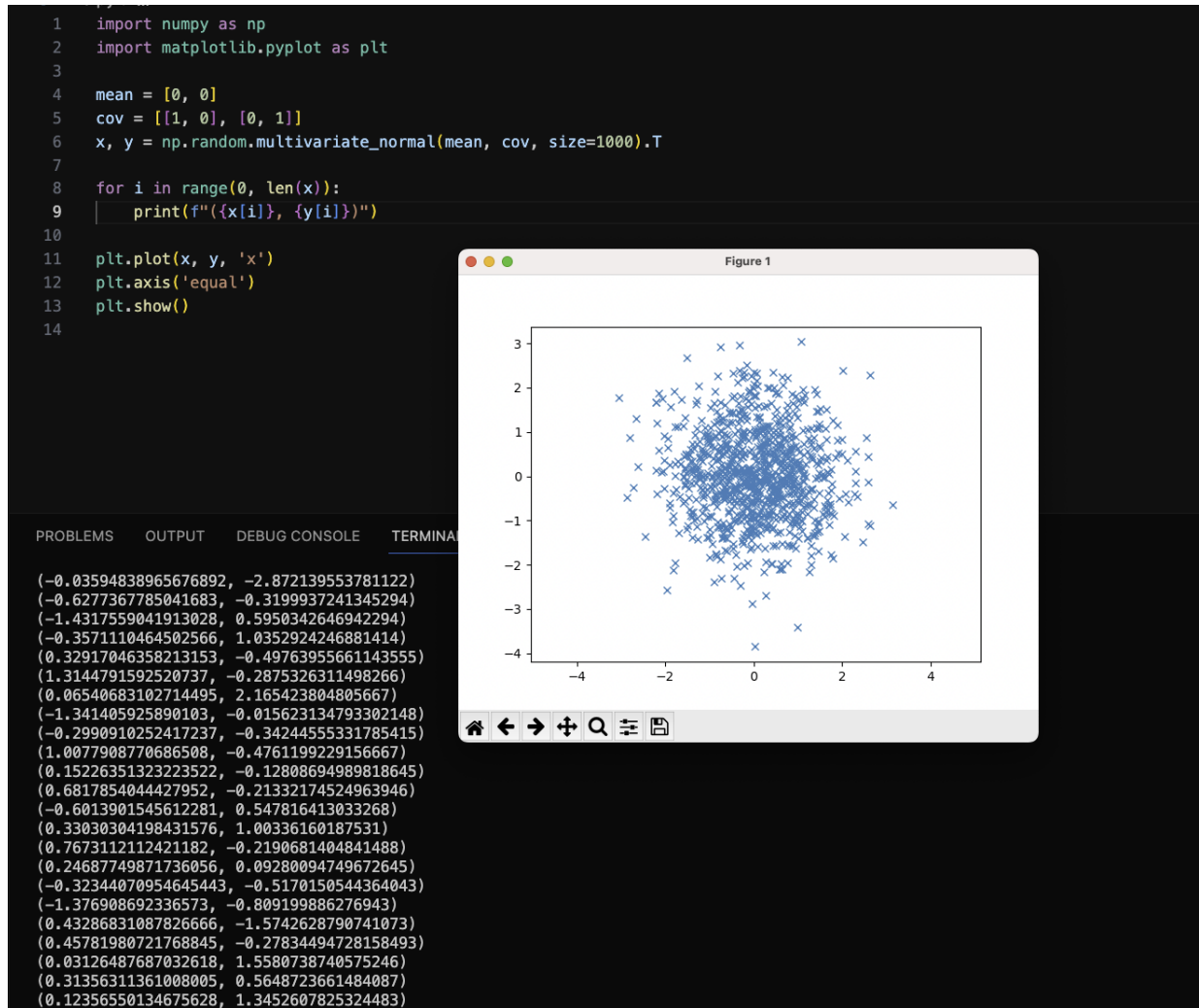
So  $f(n) \neq O(g(n))$ .

$$\lim_{n \rightarrow \infty} \frac{g(n)}{f(n)} = \lim_{n \rightarrow \infty} \frac{2^n}{3^n} = \lim_{n \rightarrow \infty} \left(\frac{2}{3}\right)^n = 0 < \infty.$$

So  $g(n) = O(f(n))$ .

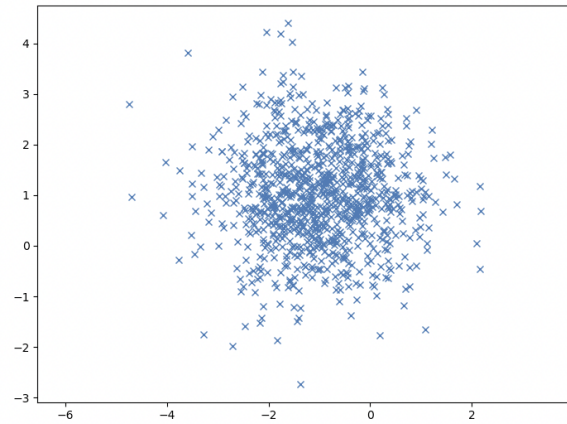
## Problem 5

- (a) Sampling from multivariate probability distributions  
i.



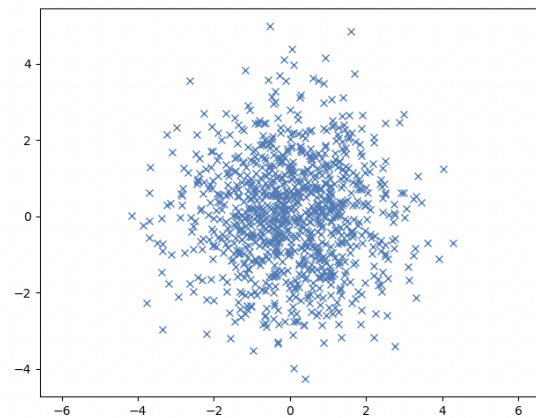
ii.

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 mean = [-1, 1]
5 cov = [[1, 0], [0, 1]]
6 x, y = np.random.multivariate_normal(mean, cov, size=1000).T
7
8 # for i in range(0, len(x)):
9 #     print(f"({x[i]}, {y[i]})")
10
11 plt.plot(x, y, 'x')
12 plt.axis('equal')
13 plt.show()
14
```



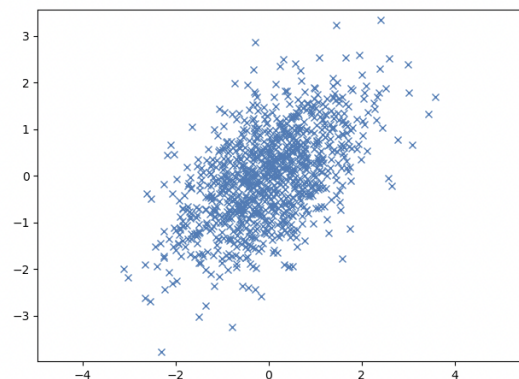
iii.

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 mean = [0, 0]
5 cov = [[2, 0], [0, 2]]
6 x, y = np.random.multivariate_normal(mean, cov, size=1000).T
7
8 # for i in range(0, len(x)):
9 #     print(f"({x[i]}, {y[i]})")
10
11 plt.plot(x, y, 'x')
12 plt.axis('equal')
13 plt.show()
14
```



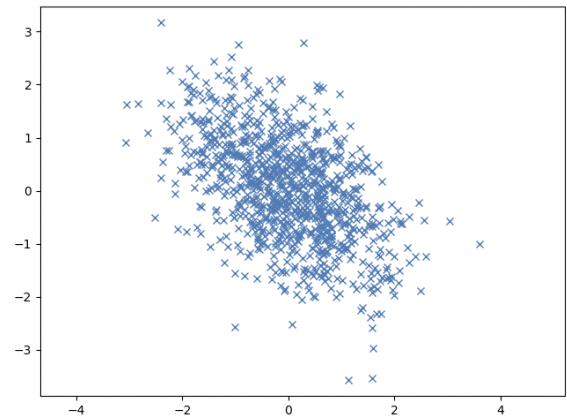
iv.

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 mean = [0, 0]
5 cov = [[1, 0.5], [0.5, 1]]
6 x, y = np.random.multivariate_normal(mean, cov, size=1000).T
7
8 # for i in range(0, len(x)):
9 #     print(f"({x[i]}, {y[i]})")
10
11 plt.plot(x, y, 'x')
12 plt.axis('equal')
13 plt.show()
14
```



V.

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 mean = [0, 0]
5 cov = [[1, -0.5], [-0.5, 1]]
6 x, y = np.random.multivariate_normal(mean, cov, size=1000).T
7
8 # for i in range(0, len(x)):
9 #     print(f"{{x[i]}}, {{y[i]}}")
10
11 plt.plot(x, y, 'x')
12 plt.axis('equal')
13 plt.show()
14
```



(b)

i.

Name: netflix\_titles.csv

Link: <https://www.kaggle.com/datasets/shivamb/netflix-shows/>

ii.

Description: this tabular dataset consists of listings of all the movies and tv shows available on Netflix, along with details such as cast, directors, ratings, release year, duration, etc.

iii.

There are 8807 examples.

iv.

There are 11 features (excluded show\_id)