اصلاحات صورت گرفته:

تصاویر نتایج حاصل شده در بخش های پیاده سازی را در گزارش قبلی نگذاشته بودم که در اینجا اضافه کردم. تمرین 5-d **و** 5-e **را جدید نوشتم. در این قسمت ها برخی نکات را قبلا بدیهی انگاشته بودم که لازم دیدم اضافه کنم.**

به نام خدا

دانشگاه صنعتی امیرکبیر

دانشکده مهندسی کامپیوتر

پاسخ تمرین سری اول شناسایی آماری الگو

استاد:

دکتر رحمتی

دانشجو:

حلیمه رحیمی

شماره دانشجویی:

۹۹۱۳۱۰۴۳

پاییز ۱۳۹۹

1. Designing Simple Pattern Analysis Systems

a. Predicting your final grade in this course	b. Predicting US Dollars to Iranian Rial exchange rate in the coming year	c. Grouping students in a dorm by their personalities
1. Regression	Regression	Clustering
2. No need	No need	No need
3. The final grade and values to the chosen features for the past terms' students	The history of the exchang rate in past years and the values to the features for those years	only the values to features for groups of students in the dorm in past years
4. The features that only students know of their value: questionares. The features related to performance during semster and the final grades: from faculty	The type of news in the websites can be collected from Internet, the history of the rate, the demand, the rise and fall in other areas can be collected from Internet as well (I guess)	The features will be collected from questionaires created by a psychologist. Also features surrounding their lifestyle gotta be through a questinaire.

a	b	c
5. The amount of time student studies, The grade for prerequisite courses, The ratio of number of times student was present in the class, The grade for each homework, grade for quizes	The amount of certain types of news, the demand for it, the rate for other money exchanges, the demand for other types of asset	Sleep cycle, Study time, discipline, The amount of time they spend on certain hobbies,
6. No need	No need	There has to be a type of rating.
7. There's a probability that some features' values would be left empty due to inability of collection	It's possible that something political happens and has a lot of effect on the rate.	There might be someone lying about their answers.
8. It would give a more exact number.	You would have a system that estimates faster than a human.	It most definitly would save a lot of time since it takes time to wait and see how each student acts

2. Getting More Familiar with the Art of Feature Extraction

a. Race Recognition (African, Non_African)
Skin color, hair texture, structure of their nose, lips, the structure of their face (their skull)

b. Facial Expression Recognition (Happy, Neutral)
The shape of the mouth, the possibility of seeing the teeth, The shape of eyes, laugh lines, dimples

c. Age Detection (Young, Adult)
Wrinckles on the forehead, Wrinckles around eyes, dark Circles around eyes, the ratio of the size of eyes to the size of face, the shape of chin, The structure of the face.

d. Gender Recognition (Male, Female)
Facial hair, the size of the jaw, the structure of the face, chin, the distance between each component of the face. in men in general and in women in general.

e. Facial Recognition
The distance between each component of the face, eyecolor, skincolor, hair Color, the structure of the face

3. Basic Statistics Warm-up

- First, find the following quantities for a random variable X with the probability density function:

$$f(x) = \begin{cases} cx & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- a. $c = 2$

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1 \Rightarrow \int_0^1 cx\,dx = c\frac{x^2}{2}\Big|_0^1 = \frac{c}{2} - 0 = 1$$

$\Rightarrow c = 2$

- b. $P(0 \leq X \leq 0.5) = \int_0^{0.5} 2x\,dx = x^2\Big|_0^{0.5} = 0.25 - 0 = 0.25$

- c. $E(x) = \int_{-\infty}^{+\infty} x f(x)\,dx = \int_0^1 x(2x)\,dx = \int_0^1 2x^2\,dx = \frac{2x^3}{3}\Big|_0^1$

$= \frac{2}{3}$

- d. $Var(x) = \Sigma(x-\mu)^2 = E(x^2) - [E(x)]^2 = \int_{-\infty}^{+\infty} x^2 f(x)\,dx - E^2(x)$

$= \int_0^1 2x^3\,dx - (\frac{2}{3})^2 = \frac{x^4}{2}\Big|_0^1 - (\frac{2}{3})^2 = \frac{1}{2} - \frac{4}{9} = \frac{9}{18} - \frac{8}{18} = \frac{1}{18}$

- e. $E[2x-2] = \int_0^1 (2x-2)\,2x\,dx = 4(\int_0^1 x^2\,dx - \int_0^1 x\,dx) = 4\left[\frac{x^3}{3} - \frac{x^2}{2}\right]_0^1$

$= \frac{4}{3} - 2 = \frac{4}{3} - \frac{6}{3} = -\frac{2}{3}$

OR

$E[2x-2] = 2E(x) - 2 = 2(\frac{2}{3}) - 2 = \frac{4}{3} - 2 = -\frac{2}{3}$

- f. $Var[2x-2] = E[(2x-2)^2] - [E(2x-2)]^2 = \int_0^1 (2x-2)^2(2x)\,dx$

$-(\frac{-2}{3})^2$ → next page

$$= \int_0^1 (8x^3 + 8x - 16x^2)\,dx - \frac{4}{9} = 8\left(\frac{x^4}{4} + \frac{x^2}{2} - \frac{2x^3}{3}\right)\Big|_0^1 - \frac{4}{9}$$

$$= 8\left(\frac{1}{4} + \frac{1}{2} - \frac{2}{3}\right) - \frac{4}{9} = 8\left(\frac{3}{4} - \frac{2}{3}\right) - \frac{4}{9} = \frac{2}{3} - \frac{4}{9} = \frac{2}{9}$$

OR

$$Var[2x-2] = 2^2\, Var(x) = 4 \times \frac{1}{18} = \frac{2}{9}$$

Now suppose a normal random variable X with parameters $\mu=1$ and $\sigma^2 = 9$.

g. Calculate $P\{-2 \leq x \leq 1\}$

$$P\{-2 \leq x \leq 1\} = P\left\{\frac{-2-1}{3} \leq \frac{x-\mu}{\sigma} \leq \frac{1-1}{3}\right\} = P\{-1 \leq z \leq 0\}$$

$$= P(z \leq 0) - P(z \leq -1) = 0.5000 - (1 - P(z \leq 1)) = 0.5000 - 0.1587 = 0.3413$$

h. Calculate $E(x)$ and $var(x)$.

$$E(x) = \int_{-\infty}^{+\infty} x\left(\frac{1}{\sigma\sqrt{2\pi}}\, e^{\left(\frac{-1}{2\sigma^2}(x-\mu)^2\right)}\right) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x\, e^{\frac{-(x-\mu)^2}{2\sigma^2}}\, dx$$

$$\overset{*}{=} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sqrt{2}\,\sigma u + \mu)(e^{-u^2})(\sqrt{2}\,\sigma)\,du$$

$$\boxed{\begin{aligned} * u &= \frac{x-\mu}{\sqrt{2}\,\sigma} \\ &\Downarrow \\ x &= \sqrt{2}\,\sigma u + \mu \\ dx &= \sqrt{2}\,\sigma\, du \end{aligned}}$$

$$= \frac{1}{\sqrt{\pi}}\left(\int_{-\infty}^{+\infty} \sqrt{2}\,\sigma u\, e^{-u^2} du + \int_{-\infty}^{+\infty} \mu\, e^{-u^2} du\right)$$

$$= \frac{1}{\sqrt{\pi}}\left[\sqrt{2}\,\sigma\left(-\frac{1}{2} e^{-u^2}\right)_{-\infty}^{+\infty} + \mu \underbrace{\int_{-\infty}^{+\infty} e^{-u^2} du}_{\text{proof: next page } *}\right]$$

$$= \frac{1}{\sqrt{\pi}}\left(0 + \mu\sqrt{\pi}\right) = \mu = 1$$

$$\ast \int_{-\infty}^{+\infty} e^{-u^2} du = \int_{-\infty}^{+\infty} e^{-t^2} dt \Rightarrow \left(\int_{-\infty}^{+\infty} e^{-u^2} du\right)^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-u^2} e^{-t^2} dt\,du$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(u^2+t^2)} dt\,du = \int_0^{2\pi}\int_0^{\infty} e^{-r^2} r\,dr\,d\theta = \int_0^{2\pi} \left[-\frac{1}{2} e^{-r^2}\right]_0^{\infty} d\theta$$

$$= \int_0^{2\pi} \frac{1}{2} d\theta = \frac{\theta}{2}\Big|_0^{2\pi} = \pi \Rightarrow \int_{-\infty}^{+\infty} e^{-u^2} du = \sqrt{\pi}$$

$$Var(x) = \sigma^2 = 9$$

$$Var(x) = \int_{-\infty}^{+\infty} x^2 \left(\frac{1}{\sigma\sqrt{2\pi}} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}\right) dx - \mu^2 \qquad \ast\ u = \frac{x-\mu}{\sqrt{2}\sigma}$$

$$= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sqrt{2}\sigma u + \mu)^2 e^{-u^2} du - \mu^2$$

$$= \frac{1}{\sqrt{\pi}}\left(2\sigma^2 \int_{-\infty}^{\infty} u^2 e^{-u^2} du + 2\sqrt{2}\sigma\mu \int_{-\infty}^{\infty} u e^{-u^2} du + \mu^2 \int_{-\infty}^{\infty} e^{-u^2} du\right) - \mu^2$$

$$= \frac{1}{\sqrt{\pi}}\left(2\sigma^2 \int_{-\infty}^{+\infty} u^2 e^{-u^2} du + 0\right) + \mu^2 - \mu^2$$

$$= \frac{2\sigma^2}{\sqrt{\pi}}\left(\left[-\frac{u}{2} e^{-u^2}\right]_{-\infty}^{+\infty} + \underbrace{\frac{1}{2}\int_{-\infty}^{+\infty} e^{-u^2} du}_{\frac{\sqrt{\pi}}{2}}\right) = \frac{2\sigma^2\sqrt{\pi}}{2\sqrt{\pi}} = \sigma^2$$

(I got to learn the proof from ProofWiki.)

- ii. Find the distribution of $Y = 2X - 1$. Express what type of random variable it is.

$P(Y=y) = P(2x-1=y) = P(x = \frac{y+1}{2})$ therefore Y is normally distributed. $E(Y) = 2E(x) - 1 = 1$ $Var(Y) = 4\,Var(x) = 36$

Then suppose that in Amirkabir University, 1/5 of the students are going to fail a certain course. Seven students are selected randomly.

j. What is the probability that exactly 4 students of them pass this course?

probability of success: p probability of failure: q

$p = \frac{4}{5}$ $q = \frac{1}{5}$

$$P(X=4) = \binom{7}{4} \left(\frac{4}{5}\right)^4 \left(\frac{1}{5}\right)^3 = \frac{7 \times \cancel{6} \times 5}{\cancel{3 \times 2 \times 1}} \times 4^4 \times \left(\frac{1}{5}\right)^7 = 7 \times 4^4 \times \left(\frac{1}{5}\right)^6$$

$$= \frac{1792}{15625} = 0.114688$$

Next, consider a continuous random variable X has the following probability density function:

$$f(x) = \begin{cases} \frac{1}{4}(4 - x^2) & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

k. Find the median value of X.

$$\int_{-\infty}^{x} f(x)\,dx = \frac{1}{2} \Rightarrow \int_{0}^{x} \frac{1}{4}(4 - x^2)\,dx = \frac{1}{2}$$

$$\left[4x - \frac{x^3}{3}\right]_0^x = 2 \Rightarrow 4x - \frac{x^3}{3} - 2 = 0 \Rightarrow x^3 - 12x + 6 = 0$$

Now, assume a dvd disc production company produces discs with a normally distributed diameters with a mean of 10 cm and standard deviation of 0.1 cm.

- 1. what is the probability of a produced disc having a diameter less than 9.8 cm?

$$P(x \leq 9.8) = P\left(\frac{x-\mu}{\sigma} \leq \frac{9.8-10}{0.1}\right) = P(z \leq 2) = 0.9772$$

Finally, assume a continuous random variable with the following probability density function:

$$f(x) = \begin{cases} \dfrac{4}{\pi(1+x^2)} & 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

calculate $E(x)$.

$$E(x) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 \frac{4x}{\pi(1+x^2)}\,dx = \frac{2}{\pi}\int_0^1 \frac{2x}{1+x^2}\,dx =$$

$$\frac{2}{\pi}\int_1^2 \frac{u'}{u}\,du = \frac{2}{\pi}\ln|u|\Big|_1^2 = \frac{2}{\pi}\ln 2 \simeq 0.69$$

4. Mastering Eigenvalues and Eigenvectors and Their Properties

In this problem, we are going to take a deeper look into these concepts.

● a. For each of the following pairs, determine whether v is an eigenvector of A or not. If so, specify the corresponding eigenvalue.

• a1. $A = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix}$, $v = \begin{bmatrix} -3 \\ -3 \end{bmatrix}$ $\qquad A\underline{v} = \lambda v$

$$\begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} -3 \\ -3 \end{bmatrix} = \begin{bmatrix} -3-12 \\ -9-6 \end{bmatrix} = \begin{bmatrix} -15 \\ -15 \end{bmatrix} = \begin{bmatrix} -3\lambda \\ -3\lambda \end{bmatrix} \Rightarrow \lambda = \frac{-15}{-3} = 5$$

• a2. $A = \begin{bmatrix} 5 & 6 & 2 \\ 0 & -1 & -8 \\ 1 & 0 & -2 \end{bmatrix}$, $v = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

$$\begin{bmatrix} 5 & 6 & 2 \\ 0 & -1 & -8 \\ 1 & 0 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5+12+2 \\ 0-2-8 \\ 1+0-2 \end{bmatrix} = \begin{bmatrix} 19 \\ -10 \\ -1 \end{bmatrix} = \begin{bmatrix} \lambda \\ 2\lambda \\ \lambda \end{bmatrix} \Rightarrow \begin{matrix} \lambda = 19 \\ 2\lambda = -10 \\ \lambda = -1 \end{matrix}$$

$\Rightarrow$ Therefore v is not an eigenvector of A.

● b. Determine a basis for the eigenspace corresponding to each of the following eigenvalues.

• b1. $A = \begin{bmatrix} 1 & -6 \\ -3 & 4 \end{bmatrix}$, $\lambda = -2$ $\qquad (A - \lambda I)x = 0$

$$\begin{bmatrix} 1+2 & -6 \\ -3 & 4+2 \end{bmatrix} x = \begin{bmatrix} 3 & -6 \\ -3 & 6 \end{bmatrix} x = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{matrix} 3x_1 - 6x_2 = 0 \\ -3x_1 + 6x_2 = 0 \end{matrix}$$

$\Rightarrow x_1 = 2x_2 \quad \Rightarrow N(A - \lambda I) = c\begin{bmatrix} 2 \\ 1 \end{bmatrix} \underset{\text{I choose}}{\Rightarrow} x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

$\Rightarrow$ next page

we want a basis, so: $x = \dfrac{1}{\sqrt{2^2+1^2}}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$

• b2. $A = \begin{bmatrix} -2 & 4 & 2 \\ 2 & 1 & -2 \\ 4 & -2 & 5 \end{bmatrix}, \lambda = 6$ $\qquad (A - \lambda I)x = 0$

$$\begin{bmatrix} -2-6 & 4 & 2 \\ 2 & 1-6 & -2 \\ 4 & -2 & 5-6 \end{bmatrix} x = \begin{bmatrix} -8 & 4 & 2 \\ 2 & -5 & -2 \\ 4 & -2 & -1 \end{bmatrix} x = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\left[\begin{array}{ccc|c} -8 & 4 & 2 & 0 \\ 2 & -5 & -2 & 0 \\ 4 & -2 & -1 & 0 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|c} -8 & 4 & 2 & 0 \\ 0 & -4 & -\frac{3}{2} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|c} -8 & 0 & \frac{1}{2} & 0 \\ 0 & -4 & -\frac{3}{2} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$\Rightarrow \begin{bmatrix} 1 & 0 & \frac{-1}{16} & 0 \\ 0 & 1 & \frac{3}{8} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \Rightarrow N(A - \lambda I) = c \begin{bmatrix} 1/16 \\ -3/8 \\ 1 \end{bmatrix}$$

$$\Rightarrow x = \sqrt{1^2 + (-6)^2 + 16^2} \begin{bmatrix} 1 \\ -6 \\ 16 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{293} \\ -6/\sqrt{293} \\ 16/\sqrt{293} \end{bmatrix}$$

• c. Calculate the eigenvalues and eigenvectors associated with the following matrices.

• c1. $A = \begin{bmatrix} 3 & 2 \\ -1 & 6 \end{bmatrix}$ $\qquad Ax = \lambda x \Rightarrow (A - \lambda I)x = 0 \Rightarrow \det(A - \lambda I) = 0$

$$\begin{vmatrix} 3-\lambda & 2 \\ -1 & 6-\lambda \end{vmatrix} = 0 \Rightarrow (3-\lambda)(6-\lambda) + 2 = 0$$

$$\Rightarrow 18 - 3\lambda - 6\lambda + \lambda^2 + 2 = 0$$

$$\Rightarrow \lambda^2 - 9\lambda + 20 = 0$$

$$\Rightarrow (\lambda - 4)(\lambda - 5) = 0 \Rightarrow \lambda_1 = 4, \lambda_2 = 5$$

$\lambda_1 = 4 \Rightarrow \begin{bmatrix} 3-4 & 2 \\ -1 & 6-4 \end{bmatrix} X_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} -1 & 2 \\ -1 & -2 \end{bmatrix} X_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\Rightarrow X_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

$\lambda_2 = 5 \Rightarrow \begin{bmatrix} 3-5 & 2 \\ -1 & 6-5 \end{bmatrix} X_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} -2 & 2 \\ -1 & 1 \end{bmatrix} X_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\Rightarrow X_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

• C2. $A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ $\quad \begin{vmatrix} -\lambda & 0 \\ 0 & -\lambda \end{vmatrix} = 0 \Rightarrow (-\lambda)^2 = 0$

$\Rightarrow \lambda^2 = 0$

$\Rightarrow \lambda_1, \lambda_2 = 0$

$\lambda_1 = 0 \Rightarrow \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} X_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

X_1, X_2 can be any non-zero vectors.

• C3. $A = \begin{bmatrix} 3 & 1 & -2 \\ 2 & 3 & -2 \\ 2 & 1 & -1 \end{bmatrix}$ $\quad \begin{vmatrix} 3-\lambda & 1 & -2 \\ 2 & 3-\lambda & -2 \\ 2 & 1 & -1-\lambda \end{vmatrix} = 0$

$\Rightarrow (3-\lambda)[(3-\lambda)(-1-\lambda)+2] - [2(-1-\lambda)+4] - 2(2-6+2\lambda) = 0$

$\Rightarrow (3-\lambda)(3-\lambda)(-1-\lambda) - 4\lambda + 12 = 0$

$\xrightarrow{\cdots} \lambda^3 - 5\lambda^2 + 7\lambda - 3 = 0 \Rightarrow (\lambda-3)(\lambda-1)^2 = 0 \Rightarrow \begin{cases} \lambda_1 = 3 \\ \lambda_2 = 1 \\ \lambda_3 = 1 \end{cases}$

$$\lambda_1 = 3 \Rightarrow \begin{bmatrix} 0 & 1 & -2 \\ 2 & 0 & -2 \\ 2 & 1 & -4 \end{bmatrix} x_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, x_1 \in N(A - \lambda_1 I)$$

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \quad \left(\text{because } \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix} = - \begin{bmatrix} -2 \\ -2 \\ -4 \end{bmatrix} \right)$$

$$\lambda_2 = 1 \Rightarrow \begin{bmatrix} 2 & 1 & -2 \\ 2 & 2 & -2 \\ 2 & 1 & -2 \end{bmatrix} x_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, x_2 \in N(A - \lambda_2 I)$$

$x_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ $\quad \lambda_3 = 1 \Rightarrow x_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ $\Rightarrow$ There isn't an independent vector.

- d. Find a 2×2 matrix A with eigenvalues $\lambda_1 = -1$ and $\lambda_2 = 7$, and corresponding eigenvectors $v_1 = \begin{bmatrix} -3/2 \\ 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}$.

v_1 and v_2 are independent.

A is diagonalizable $\Rightarrow A = V \Lambda V^{-1}$

$$V = \begin{bmatrix} -3/2 & 1/2 \\ 1 & 1 \end{bmatrix} \quad \Lambda = \begin{bmatrix} -1 & 0 \\ 0 & 7 \end{bmatrix} \quad V^{-1} = \frac{1}{-\frac{3}{2} - \frac{1}{2}} \begin{bmatrix} 1 & -\frac{1}{2} \\ -1 & -\frac{3}{2} \end{bmatrix}$$

$$\Rightarrow V^{-1} = \begin{bmatrix} -\frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{4} \end{bmatrix} \quad A = \begin{bmatrix} -\frac{3}{2} & \frac{1}{2} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 7 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{4} \end{bmatrix}$$

$$A = \begin{bmatrix} \frac{3}{2} & \frac{7}{2} \\ -1 & 7 \end{bmatrix} \begin{bmatrix} \frac{-1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{4} \end{bmatrix} = \begin{bmatrix} -\frac{3}{4} + \frac{7}{4} & \frac{3}{8} + \frac{21}{8} \\ \frac{1}{2} + \frac{7}{2} & \frac{-1}{4} + \frac{21}{4} \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix}$$

- e. Let A be an invertible matrix with eigenvalue λ. Assuming a non-zero x satisfies $Ax = \lambda x$ show that A^{-1} has an eigenvalue equal to λ^{-1}.

Since A is invertible, $\lambda \neq 0$.

$$Ax = \lambda x \Rightarrow A^{-1} A x = A^{-1} \lambda x$$

$$\Rightarrow \quad x = \lambda A^{-1} x \Rightarrow A^{-1} x = \left(\frac{1}{\lambda}\right) x$$

Therefore $\frac{1}{\lambda}$ is an eigenvalue of A^{-1}.

- f. Assume A is a matrix of the size 3×3 with two eigenvalues, and the corresponding eigenspaces are one-dimensional. Justify whether A is diagonizable or not.

Because here we only have two independent eigenvectors, the matrix composed of eigenvectors is not gonna be invertible. Therefore you can't diagonalize A.

Also you could say not enough eigenvectors to have $\mathbb{R}^3$ eigenspace.

- g. Show that if an $n \times n$ matrix A has an eigenvalue equal to 3, then A^2 has an eigenvalue equal to 9.

If A is diagonalizable then $A = V \Lambda V^{-1} \Rightarrow A^2 = V \Lambda \underbrace{V^{-1} V}_{I} \Lambda V^{-1}$

$\Rightarrow A^2 = V \Lambda^2 V^{-1}$ $\Rightarrow$ since Λ is a diagonal matrix

$\Rightarrow A^2$ is gonna have an eigenvalue equal to λ^2.

• h. Consider the sequences x_n and y_n, such that for each $n \geq 1$, $x_n = 2x_{n-1} - 3y_{n-1}$, $y_n = -4x_{n-1} + y_{n-1}$

Assuming $x_0 = 2$ and $y_0 = 3$, specify each of x_n and y_n explicitly in terms of n.

$$u_n = \begin{bmatrix} 2 & -3 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}_n \qquad u_0 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\begin{vmatrix} 2-\lambda & -3 \\ -4 & 1-\lambda \end{vmatrix} = (2-\lambda)(1-\lambda) - 12 = 0 \Rightarrow 2 - 2\lambda - \lambda + \lambda^2 - 12 = 0$$

$$\Rightarrow \lambda^2 - 3\lambda - 10 = 0 \Rightarrow (\lambda - 5)(\lambda + 2) = 0 \Rightarrow \begin{cases} \lambda_1 = 5 \\ \lambda_2 = -2 \end{cases}$$

$$\lambda_1 = 5 \Rightarrow \begin{bmatrix} -3 & -3 \\ -4 & -4 \end{bmatrix} x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow x_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\lambda_2 = -2 \Rightarrow \begin{bmatrix} 4 & -3 \\ -4 & 3 \end{bmatrix} x_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow x_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

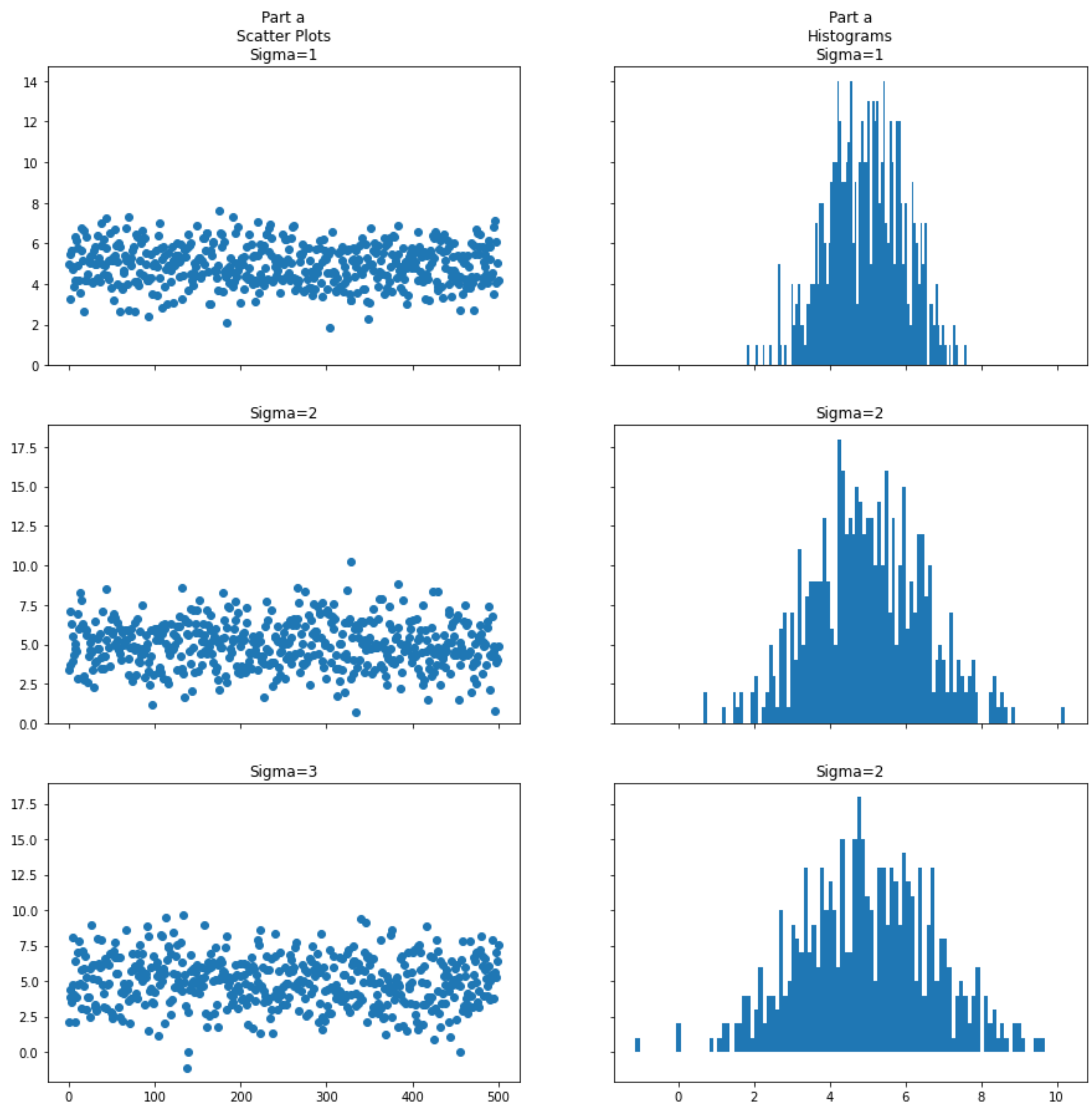
$$u_n = c_1 (5)^n \begin{bmatrix} 1 \\ -1 \end{bmatrix} + c_2 (-2)^n \begin{bmatrix} 3 \\ 4 \end{bmatrix} \qquad u_0 = c_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + c_2 \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\Rightarrow \left[\begin{array}{cc|c} 1 & 3 & 2 \\ -1 & 4 & 3 \end{array}\right] \Rightarrow \left[\begin{array}{cc|c} 1 & 3 & 2 \\ 0 & 7 & 5 \end{array}\right] \Rightarrow \begin{cases} c_1 + 3c_2 = 2 \\ 7c_2 = 5 \end{cases} \Rightarrow \begin{cases} c_1 = \frac{-1}{7} \\ c_2 = \frac{5}{7} \end{cases}$$

$$\Rightarrow x_n = \left(-\tfrac{1}{7}\right)(5)^n + \left(\tfrac{15}{7}\right)(-2)^n, \quad y_n = \left(\tfrac{1}{7}\right)(5)^n + \left(\tfrac{20}{7}\right)(-2)^n$$

5. Simple Sample Generation and Beyond

a. Generate samples from three normal distributions specified by the following parameters:
$n = 1, \quad N = 500, \quad \mu = 5, \quad \sigma = 1, 2, 3$
Plot the samples, as well as the histograms associated with each of the distributions. Compare the results.

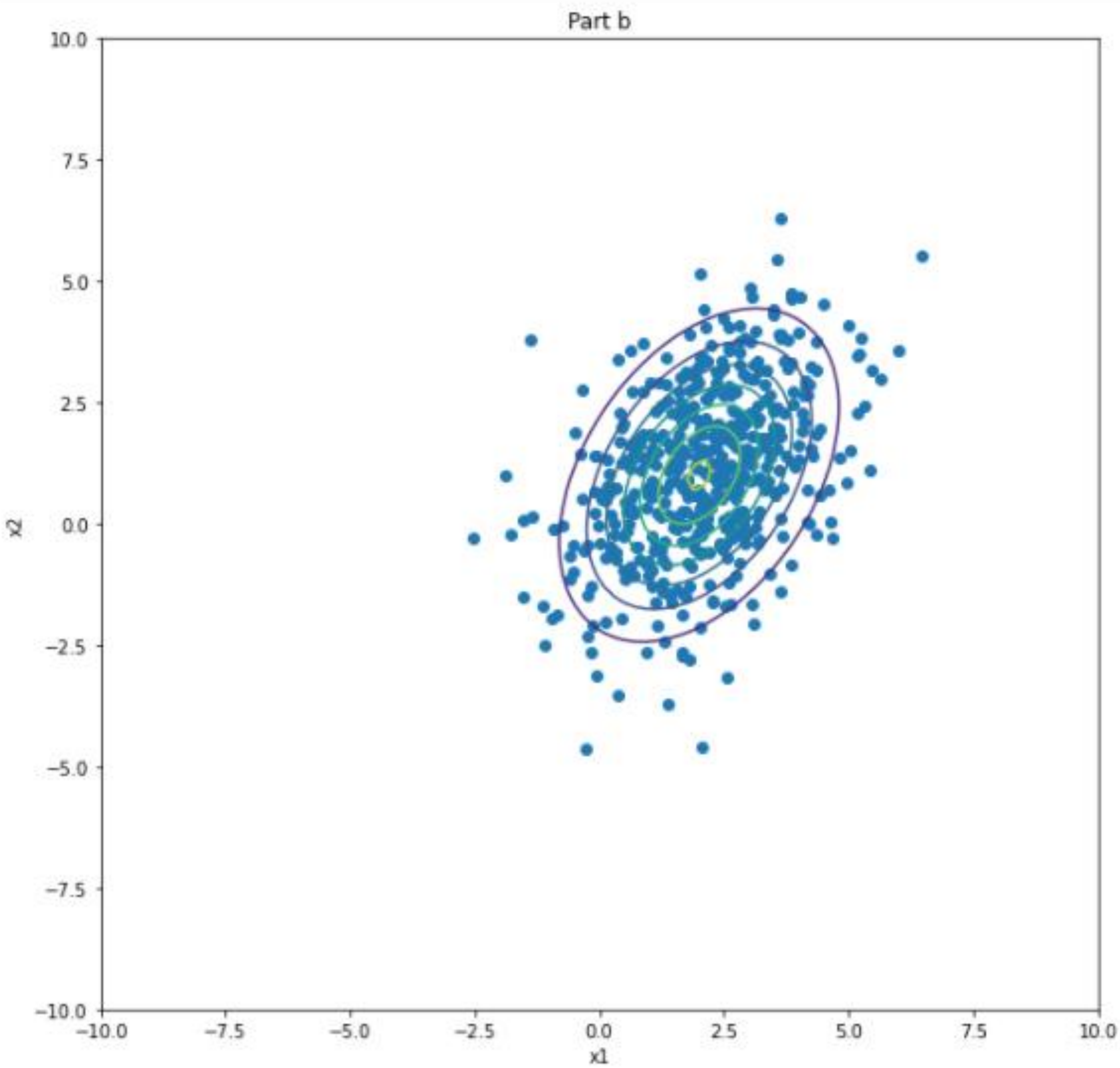


Since all three sample sets have μ=5, in the scatter plot the density of samples would be greater around 5. Also since variance increases from first to last sample set, samples' distribution would take up more vast area. The histogram for samples would slowly get to have a greater width and smaller height.

b. Generate samples from a normal distributions specified by the following parameters:

$$n = 2, \quad N = 500, \quad M = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

Display the samples, as well as the associated contour plot.



c. Consider a normal distribution specified by the following parameters:

$$n = 2, \quad N = 500, \quad M = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Determine appropriate values for each of the unknown variables, so that the shape of the distribution becomes:

c1. A circle in the upper left of the Euclidean coordinate system.

c2. A diagonal line (/ shape) in the centre

c3. A horizontal ellipsoid in the lower right of the Euclidean coordinate system

Display the generated samples.

In order to have horizontal or vertical ellipsoids (including circles) you need to consider the variables independent of one another; meaning consider correlation coefficient equal to zero.

If Circle: Consider the diagonal values equal.

If Horizontal Ellipsoid: Consider the first variable's variance greater than the other.

In order to have a diagonal distribution, choose covariance matrix as below:

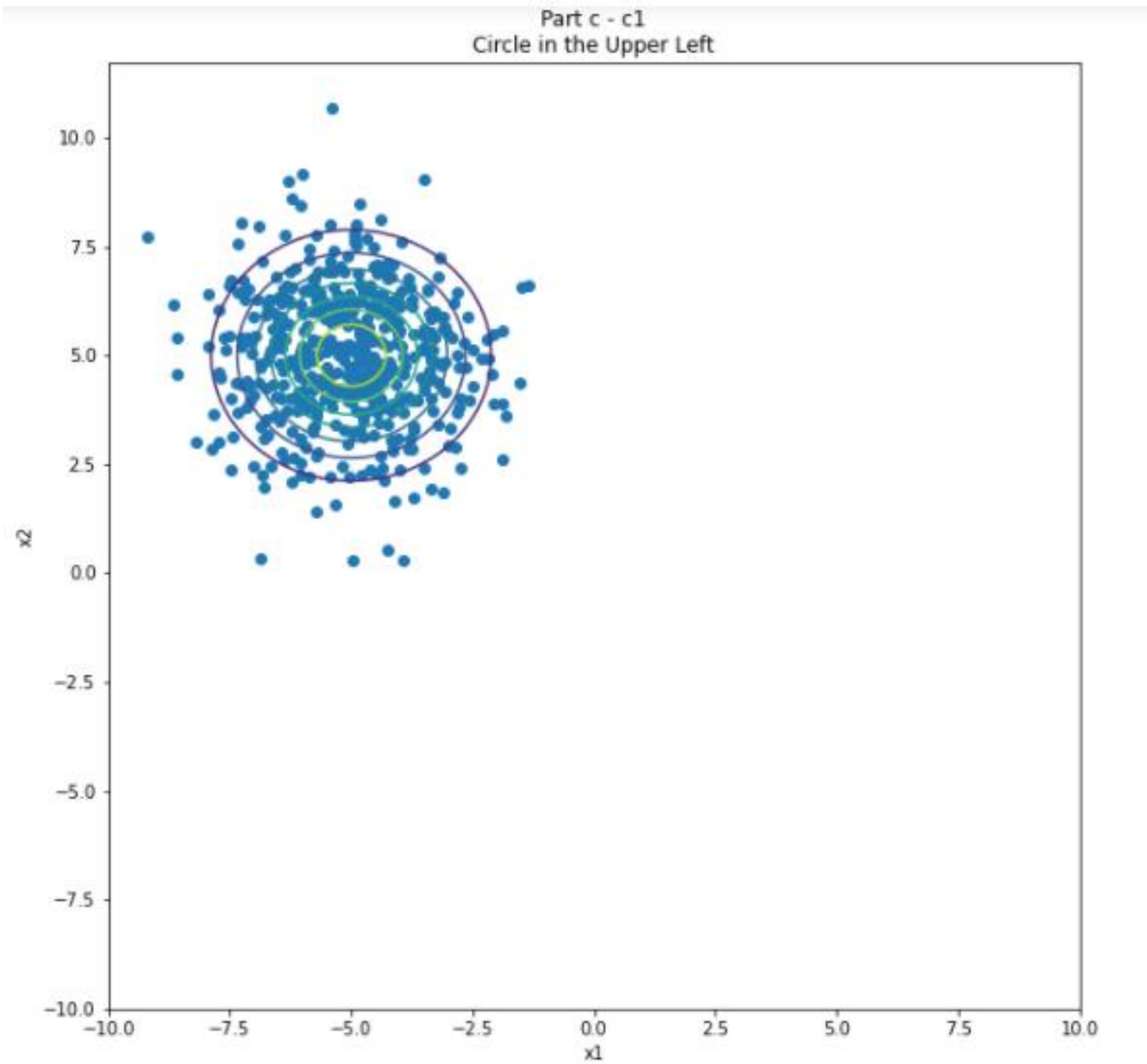
$$\sum = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 - \varepsilon \\ \rho\sigma_1\sigma_2 - \varepsilon & \sigma_1^2 \end{bmatrix}$$

I subtract ε value so the matrix doesn't become singular.

I chose the means in a way that the distributions end up in the place asked in the question.

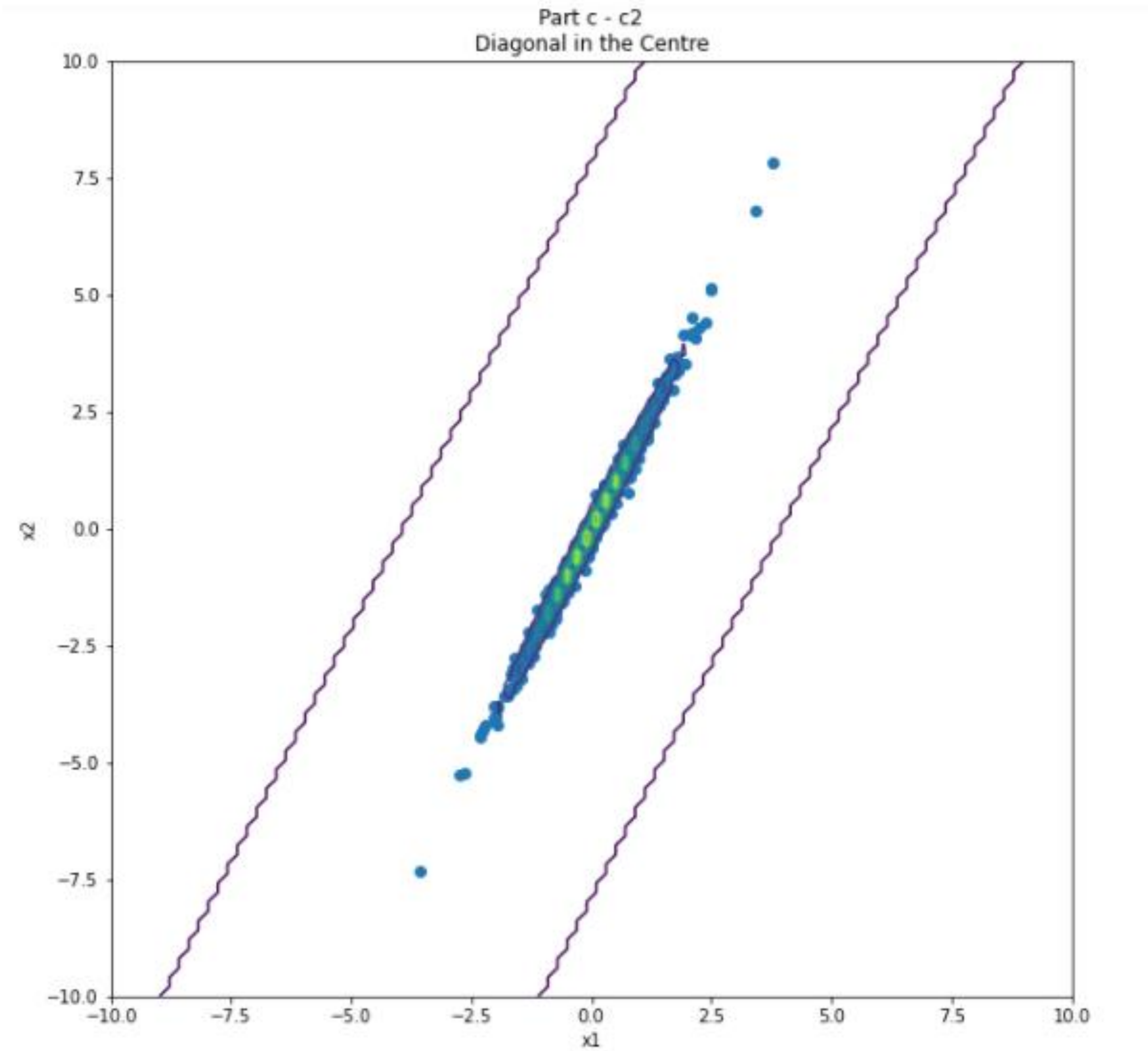
C1.

$M = \begin{bmatrix} -5 \\ 5 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



C2.

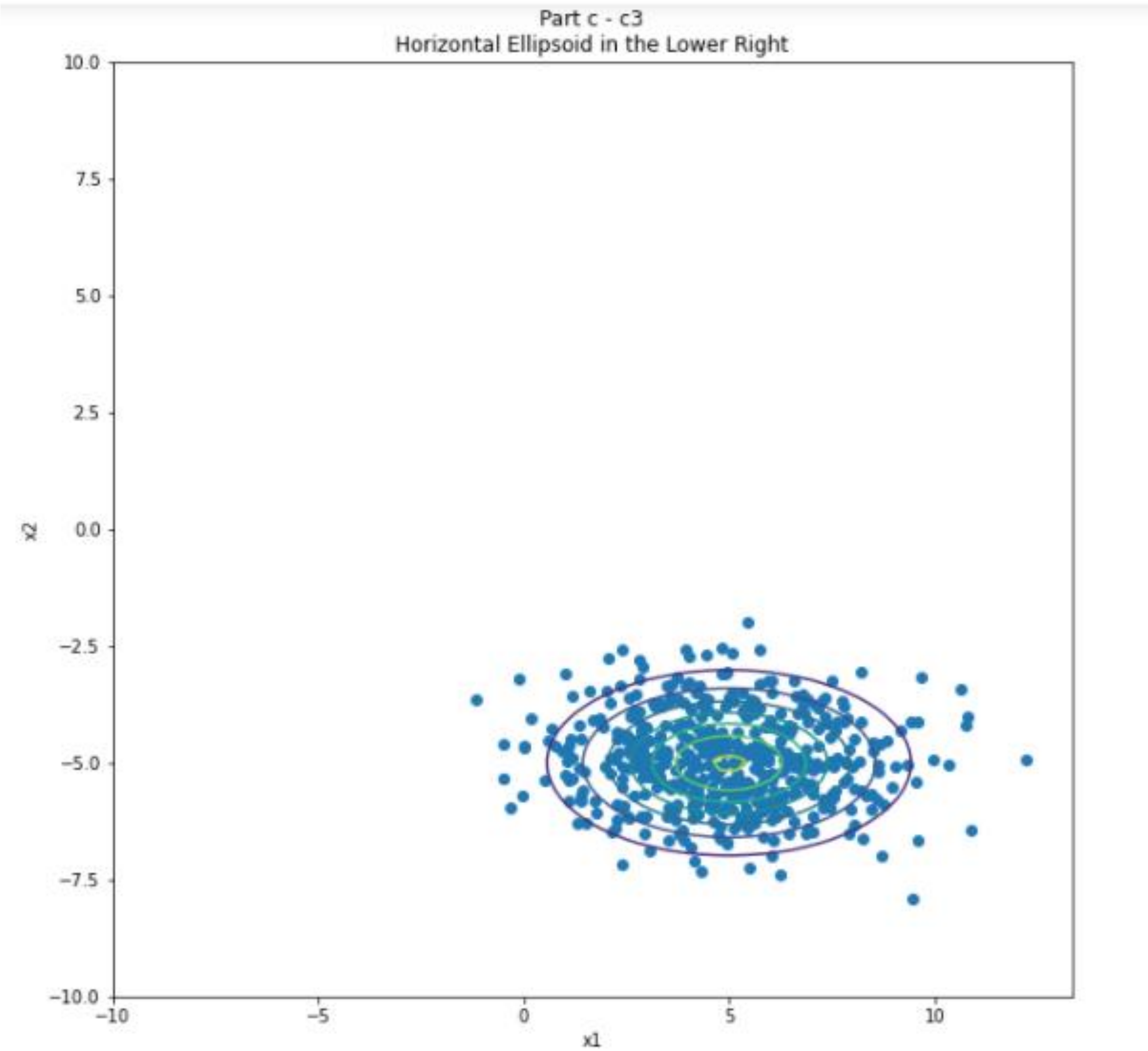
$$M = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 1 & 1.99 \\ 1.99 & 4 \end{bmatrix}$$



(The two lines are contour lines, just ignore)

C3.

$M = \begin{bmatrix} 5 \\ -5 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$



d. Consider a random variable with

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad M = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Compute $d^2(X)$ analytically, if the parameters are selected as

$$m_1 = 2, m_2 = 3, \sigma_1^2 = 1, \sigma = 4$$
$$\rho = -0.99, -0.5, 0.5, 0.99$$

$$d^2(x) = (x-\mu)^T \Sigma^{-1}(x-\mu)$$

$$= [x_1 - m_1 \quad x_2 - m_2] \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{bmatrix}\begin{bmatrix} x_1-m_1\\ x_2-m_2\end{bmatrix}$$

$$= \frac{1}{4(1-\rho^2)}\left[(x_1-m_1)\sigma_2^2 - (x_2-m_2)\rho\sigma_1\sigma_2 \quad (x_2-m_2)\sigma_1^2 - (x_1-m_1)\rho\sigma_1\sigma_2\right]$$

$$\begin{bmatrix}x_1-m_1\\x_2-m_2\end{bmatrix} = \frac{1}{4(1-\rho^2)}\Big[(x_1-m_1)^2\sigma_2^2 - (x_1-m_1)(x_2-m_2)\rho\sigma_1\sigma_2$$

$$+ (x_2-m_2)^2\sigma_1^2 - (x_1-m_1)(x_2-m_2)\rho\sigma_1\sigma_2\Big]$$

$$= \frac{1}{4(1-\rho^2)}\left[(x_1-2)^2\, 4 - 4(x_1-2)(x_2-3)\,\rho + (x_2-3)^2\right]$$

$$= \frac{1}{4(1-\rho^2)}\left[(2x_1-4)^2 + (x_2-3)^2 - 4\rho(x_1-2)(x_2-3)\right]$$

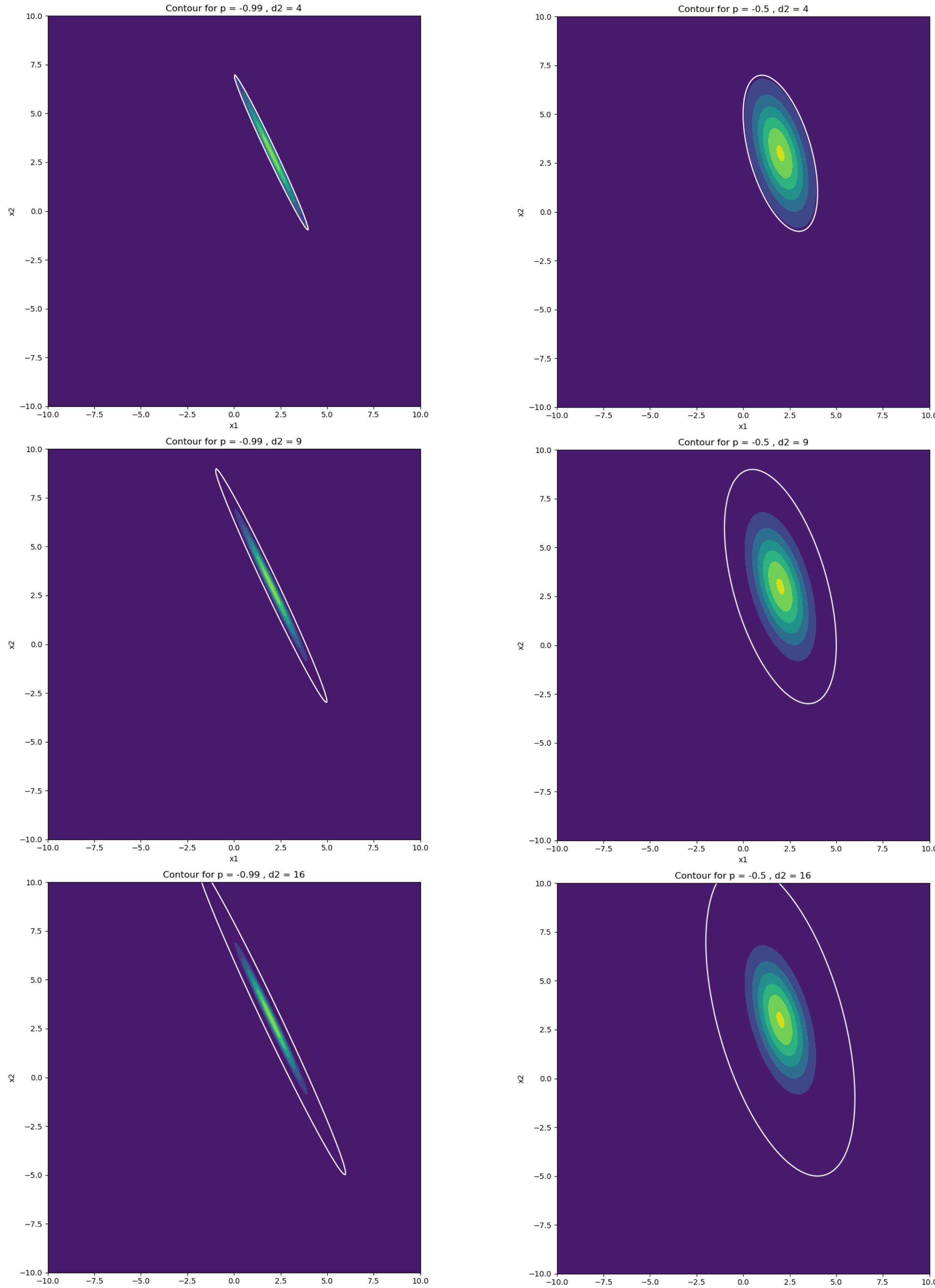
d.

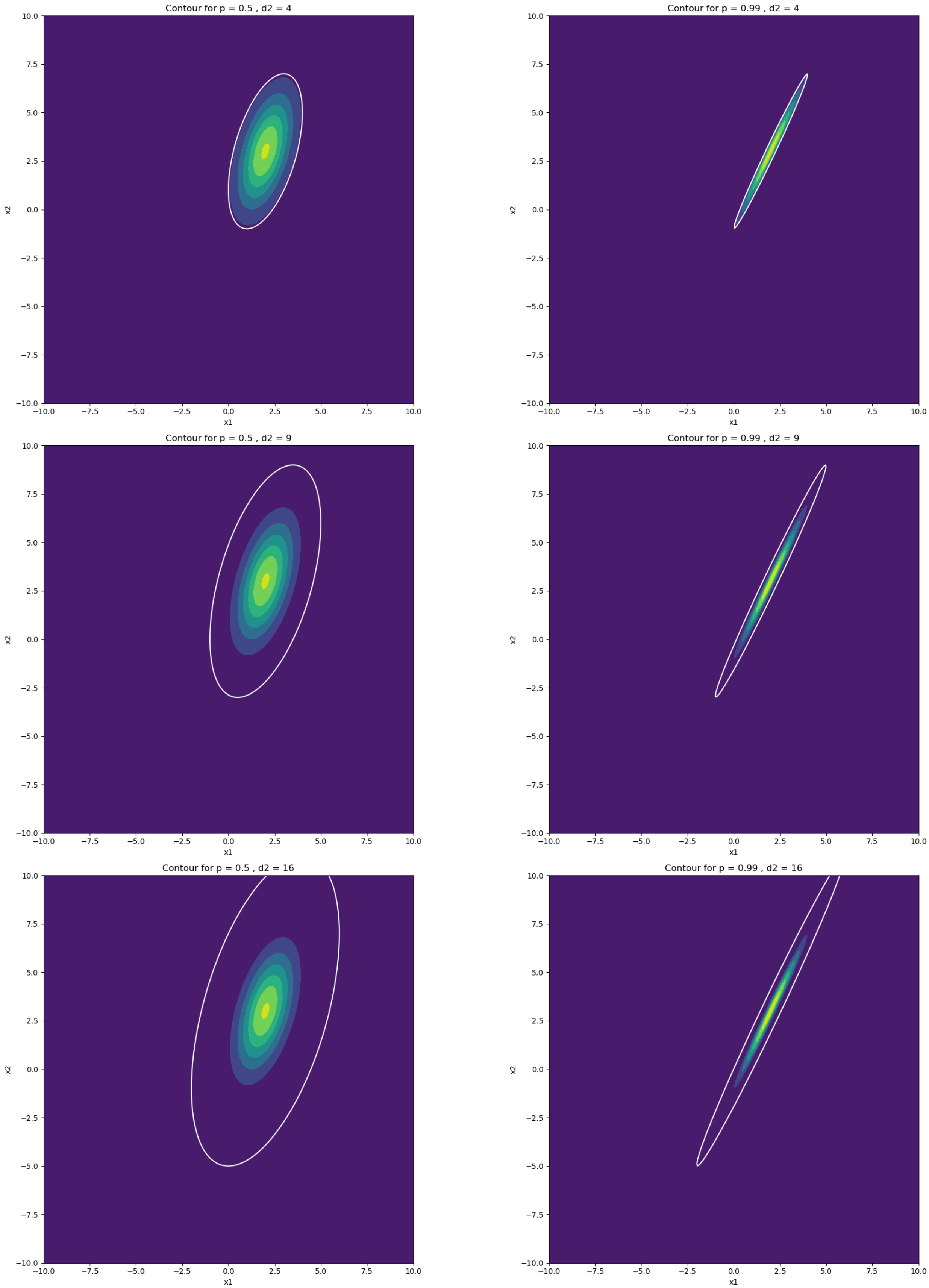
```
ro = [-0.99, -0.5, 0.5, 0.99]
```

```
#Printing the Equations
for p in ro:
    print("d^2 = %s  ((2*y-4)^2 + (x-3)^2 - 4(%s)(y-2)(x-3))" % (1/(4*(1-p**2)),p))
```

```
d^2 = 12.56281407035174  ((2*y-4)^2 + (x-3)^2 - 4(-0.99)(y-2)(x-3))
d^2 = 0.3333333333333333  ((2*y-4)^2 + (x-3)^2 - 4(-0.5)(y-2)(x-3))
d^2 = 0.3333333333333333  ((2*y-4)^2 + (x-3)^2 - 4(0.5)(y-2)(x-3))
d^2 = 12.56281407035174  ((2*y-4)^2 + (x-3)^2 - 4(0.99)(y-2)(x-3))
```

e. Plot the contour lines for $d^2(X) = 4, 9, 16$.


Contour for p = -0.99 , d2 = 4
Contour for p = -0.5 , d2 = 4
Contour for p = -0.99 , d2 = 9
Contour for p = -0.5 , d2 = 9
Contour for p = -0.99 , d2 = 16
Contour for p = -0.5 , d2 = 16
x1
x2
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0


Contour for p = 0.5 , d2 = 4
Contour for p = 0.99 , d2 = 4
Contour for p = 0.5 , d2 = 9
Contour for p = 0.99 , d2 = 9
Contour for p = 0.5 , d2 = 16
Contour for p = 0.99 , d2 = 16
x1
x2

f. Calculate the sample mean $\hat{M}$, and sample covariance matrix $\hat{\Sigma}$ of the distribution in part b., and comment on the results.

The sample mean and covariance matrix would be slightly different from the mean and covariance matrix of the society (here: given normal distribution values). If the number of samples increases then sample mean and covariance will get closer to those of the society.

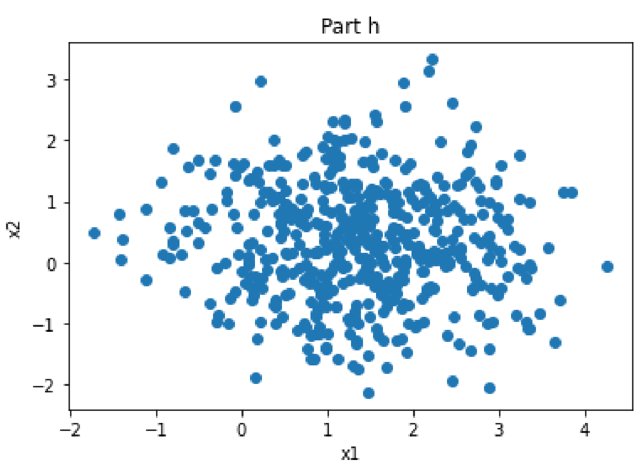
```
Part f
Mean:
 [2.08120435 1.10044719]
Covariance:
 [[1.95045833 1.00042881]
 [1.00042881 2.9171646 ]]
```

g. Simultaneously diagonalise Σ and $\hat{\Sigma}$, and form a vector $V = [\lambda_1, \lambda_2]^T$.

I assumed that by simultaneously you meant to put them in one matrix and compute the eigenvalues and eigenvectors. The result is shown below:

```
Part g
V:
 [[1.38196601 1.32273629]
 [3.61803399 3.54488665]]
```

h. Find a transformation for covariance matrix of the distribution in part b., such that when applied on the data, the covariance matrix of the transformed data becomes $\mathbf{I}$. Transform the data and display the distribution in the new space.




```
Part h
Transformation Matrix:
 [[ 0.76084521 -0.14530851]
 [-0.14530851  0.61553671]]
New Covariance:
 [[0.9694772  0.01309609]
 [0.01309609 0.96749212]]
```

We can use $d^2(x) = (x-\mu)^T \Sigma^{-1} (x-\mu)$

$$\Rightarrow d^2(x) = \left[(x-\mu)^T \Sigma^{-\frac{1}{2}}\right]\left[\Sigma^{-\frac{1}{2}} (x-\mu)\right]$$

$$= \left[\Sigma^{-\frac{1}{2}} (x-\mu)^T\right]^T \left[\Sigma^{-1} (x-\mu)\right] = y^T y = \|y\|$$

We want to transform Σ to I, so consider A is a transformation matrix

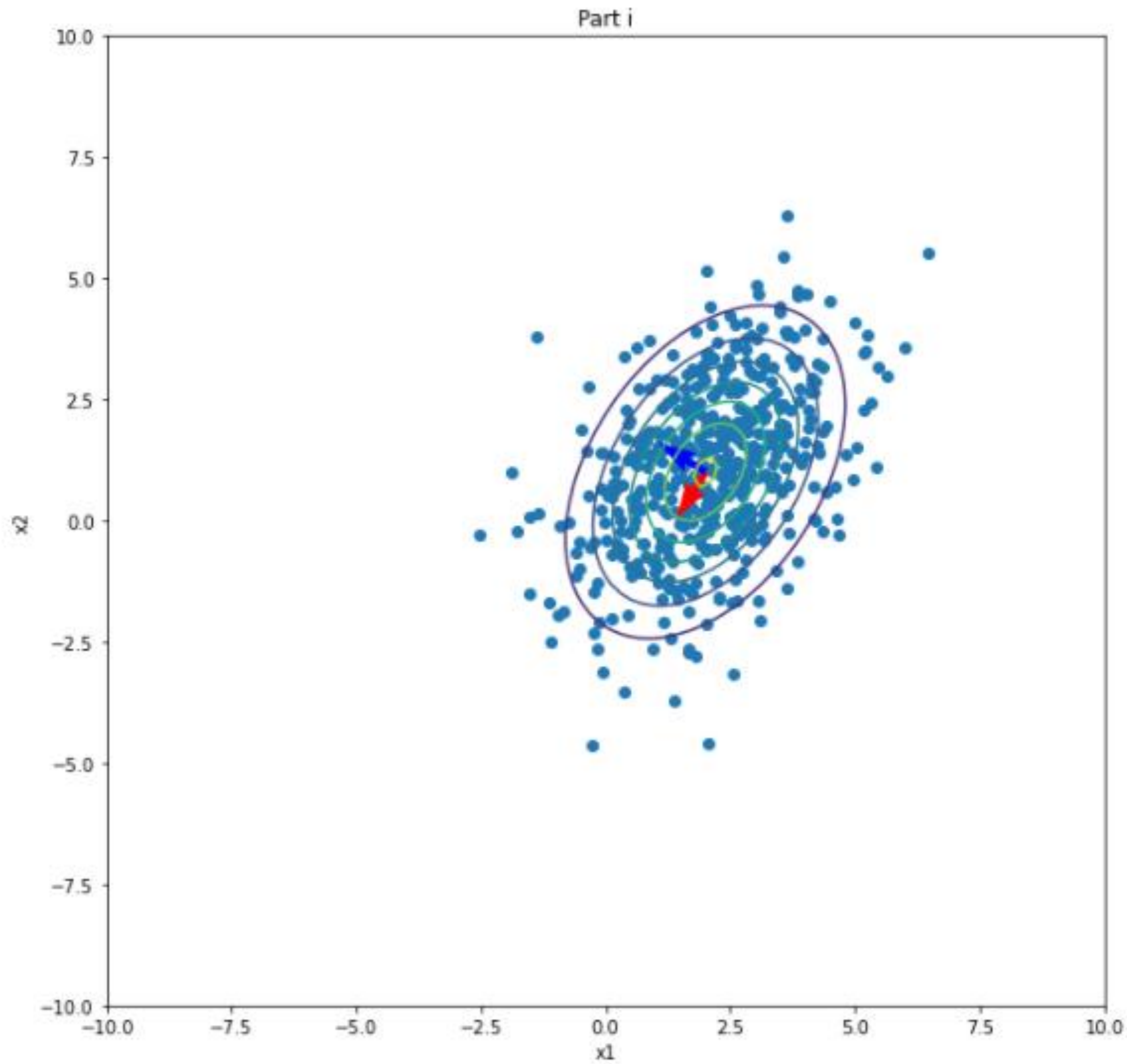
$$\Rightarrow \Sigma_{new} = A \Sigma A^T = A(\Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}}) A^T$$

if $A = \Sigma^{-\frac{1}{2}} \Rightarrow (\Sigma^{-\frac{1}{2}} \Sigma^{\frac{1}{2}})(\Sigma^{\frac{1}{2}} \Sigma^{-\frac{1}{2}}) = I$

i. Calculate the eigenvalues and eigenvectors associated with the covariance matrix of the distribution in part b. Plot the eigenvectors. What can you infer from them?

```
Part i
Eigenvalues of Distribution Part b:
 [1.38196601 3.61803399]
Eigenvectors of Distribution Part b:
 [[-0.85065081 -0.52573111]
 [ 0.52573111 -0.85065081]]
```

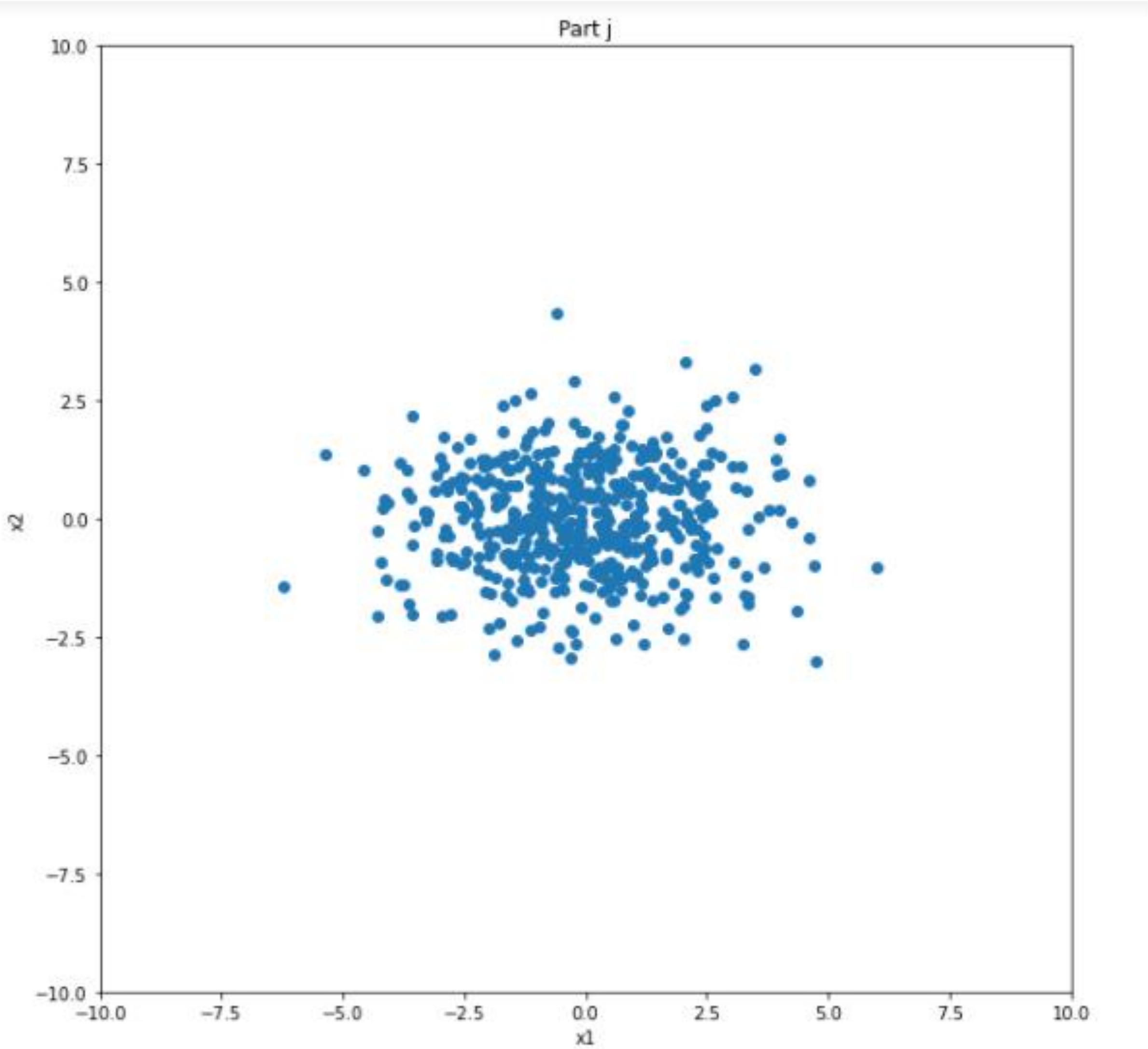
The direction of eigenvectors will be aligned with the distribution’s largest variances. Here there are two variables and we end up having two eigenvectors.



j. Again, consider the distribution and samples you generated in part b. Construct a 2×2 matrix $\mathbf{P}$, which has eigenvectors associated with Σ as its columns ($\mathbf{P} = [v_1, v_2]$, such that v_1 is corresponding to the largest eigenvalue). Project your generated samples to a new space using $\mathbf{Y}_i = (\mathbf{X}_i - M) \times \mathbf{P}$, and plot the samples. What differences do you notice?

```
Part j
P:
[[-0.52573111 -0.85065081]
 [-0.85065081  0.52573111]]
```

Because of $X - \mu$ all the samples gather around $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and since we are using the eigenvectors to project our samples to a new space, there wouldn't be any correlations between the two coordinate quantities of the samples. In other words if we consider each coordinate quantity of the samples as a value of a feature, this projection would give us a new pair of features that are independent of one another.



k. Find the covariance matrix associated with the projected samples in part h. Also calculate its eigenvalues and eigenvectors, and comment on the results.

```
Part k
New Samples Covariance:
 [[3.54478429 0.01508118]
 [0.01508118 1.32283865]]
New Samples Eigenvalues:
 [3.54488665 1.32273629]
New Samples Eigenvectors:
 [[ 0.99997697 -0.00678691]
 [ 0.00678691  0.99997697]]
```

The covariance matrix would be nearly equal to $\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$. I guess if the amount of samples increases it would be exactly equal to the above matrix.

The eigenvalues are equal to those of the covariance matrix of the samples (not the society).

The eigenvectors from the columns of an Identity matrix, since the new features would be independent of one another.

6. Some Explanatory Questions

a) Why do you think Centeral Limit Theorem is important? Where and how can it be used?

Sometimes we don't have the characteristics of the population our sample is from. Sometimes we have information on a population but not about its samples. Eigther way Centeral Limit Theorem comes in handy.

It can be used to know if a sample is from a certain population. It can also be used to know the characteristics of the samples of a population we do a similar thing in Bayes classifiers, so this is probably one of its uses.

b) What is the difference between a feature and a measurement?

Measurment is a numerical quantitation of an attribute of an object while a feature is an attribute. What we consider as a feature in pattern recognition is a charectristic of the data we are working with. A feature itself gets measured.

c) Does a covariance matrix need to be symmetric? Why?

Yes, Because covariance matrix is computed as:

$$Cov(x) = E\left[(x-\mu_x)(x-\mu_x)^T\right] = E(xx^T) - \mu_x \mu_x^T$$

As we know from Algebra: AA^T is a symmetric matrix. and the subtraction of two symmetric matrices is also a symmetric matrix.

Since $E(x_1 x_2)$ for the same $f(x)$ is equal to $E(x_2 x_1)$ therefore $E(xx^T)$ would be symmetric.

$\Rightarrow$ This means covariance matrix has to be symmetric.

d) What does zero eigenvalue mean?

Zero eigenvalue means that the matrix is singular

$$\det(A) = \prod_{i=1}^{n} \lambda_i$$

if : $\lambda = 0 \Rightarrow \det(A) = 0 \Rightarrow A$ is a singular matrix $\Rightarrow A$ is noninvertible.

$Ax = 0 \cdot x = 0 \Rightarrow$ This means there might exist a non-zero x (if A is not a zero matrix).

e) Wha does the whitening transformation come into use?

Besed on what whitening transformation does to the data, which is, it both scales and makes the features independant (feature scaling and decorrelation), I would

say it comes in handy in feature selection. We use eigenvalues and eigenvectors for the same reason (decorrelation). This way the new features are uncorrelated. Since we also have feature scaling I'm sure it makes learning faster.
It probably gets used for preprocessing images as well because we would have a lot of pixels with correlated values near eachother.