CS-460. Assignment-1

Name- Suraj Kumar Patel Roll no.-181163

Vector & Matrices

1. $y.z \rightarrow y^T z$

$y = [1, 3]$ $z = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

$y.z =$ ~~...~~ $2+9=11$

2. $Xy = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2+12 \\ 1+9 \end{bmatrix} = \begin{bmatrix} 14 \\ 10 \end{bmatrix} = Xy$

3. If $\det(X) \neq 0$ X is invertible

$\det(x) = 6-4 = 2.$

$X^{-1} = \frac{\operatorname{adj}(X)}{\det|X|}$

$\operatorname{adj}(X) = \begin{pmatrix} 3 & -4 \\ -1 & 2 \end{pmatrix}$, $\det|X| = 2$

$X^{-1} = \frac{1}{2}\begin{pmatrix} 3 & -4 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 3/2 & -2 \\ -1/2 & 1 \end{pmatrix}$

4. Matrix is 2×2 and $\det|X| \neq 0$ so rank is 2

Calculus

1.

$$y = x^3 + x - 5$$

$$\frac{dxy}{dx} = 3x^2 + 1$$

2. $$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix}$$

$$f(x_1, x_2) = x_1 \sin(x_2) e^{-x_1}$$

$$\frac{\partial f}{\partial x_1} = \cancel{x_1 e^{-x_1}} \sin(x_2) \frac{\partial}{\partial x_1}(x_1 e^{-x_1})$$

$$= \sin(x_2)\left[e^{-x_1} - x_1^2 e^{-x_1}\right]$$

$$\frac{\partial f}{\partial x_1} = e^{-x_1}\sin(x_2) - \sin(x_2) x_1^2 e^{-x_1}$$

$$\frac{\partial f}{\partial x_2} = x_1 \cos(x_2) e^{-x_1}$$

$$\nabla f(x) = \begin{pmatrix} \sin(x_2)\left[e^{-x_1} - x_1^2 e^{-x_1}\right] \\ x_1 \cos(x_2) e^{-x_1} \end{pmatrix}$$

Probability

$S = \{1, 1, 0, 1, 0\}$

1) Sample mean

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{X} = \frac{1+1+0+1+0}{5} = \frac{3}{5}$$

2) Sample variance

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$$= \frac{\left(1-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2 - \frac{3}{5}^2 + \left(1-\frac{3}{5}\right)^2 - \frac{3}{5}^2}{5-1}$$

$$S^2 = \cancel{3.0}\ 0.3$$

3) probility $= \frac{1}{2^5}$

4) Let $P(X=1) = P$ so $P(X=0) = 1-P$

Probability of sample space. S.

$$P(x) \cdot \quad p^3(1-p)^2 = p^3(1-p)^2$$

$$\frac{dP(}{dp} = 3p^2(1-p)^2 - 2p^3(1-p)$$

5. (a) $P(Z=T \text{ AND } y=b) = 0.1$

(b) $P(Z=T \mid y=b)$

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

$$\Rightarrow P(Z=T \mid y=b) = \frac{P(Z=T \cap y=b)}{P(B)}$$

$$= \frac{0.1}{0.25} = \frac{10}{25} = \frac{2}{5}$$

<u>Big O</u>

for $f(n) = O(g(n))$ $\exists$ a positive integer c and n_0 s.c. $\forall n \geq n_0$

$$0 \leq f(n) \leq c \cdot g(n)$$

1. $f(n) = \ln(n)$

$g(n) = \log_2(n)$

$$\log_2(n) = \frac{\log_e n}{\log_e 2} \therefore$$

$$\Rightarrow \lg(n) = \frac{\ln(n)}{\ln(2)}$$

So we can ~~write~~ say they are dependent so for different value of c we can get

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

2. $f(n) = 3^n$

$g(n) = n^{100}$

lets $n_0 = 1000$ then for $c = 1$ $\forall n \geq 100$

$\therefore f(n) > g(n)$ as $3^{1000} < 1000^{100}$

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

3. $f(n) = 3^n$

$g(n) = 2^n$

since $3 > 2$ for any n_0 and $c = 1$

$0 < g(n) < f(n)$

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

4. $f(n) = 1000n^2 + 2000n + 4000$

$g(n) = 3n^2 + 1$

for very larg n_0 and for $C = 1$ $\forall n \geq n_0$

$0 < f(n) \leq g(n)$

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

Algorithms

Devide and Conquer

To find

arr : $\underbrace{0\ 0\ 0\ 0\ 0\ .\ 01.\ 1\ 1\ 1\ 1\ 1\ 1}_{n}$

as all 0's appear before 1. first will devide the array in half and check the middle element and the next element, if both are zeroes then the transition point is on the right, if both are ones then the trasition point is on left.

→ Then delete the left/right part according to that and continue the process till reach a point where the mid is zero and next to mid is one

→ we are only search half the array so runtime is $O(\frac{n}{2})$

By masters theorem runtime is $O(\log n)$.

Probabillity and Random variables

(a) False, as $A \cup B \neq \phi$ and $A \cap (B \cap A') = \phi$

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

(c) $P(A) = P(A \cap B) + P(A' \cap B)$; False

d) $P(A|B) = P(B|A)$ False

$P(A|B) = \frac{P(A \cap B)}{P(B)}$, $P(B|A) = \frac{P(A \cap B)}{P(A)}$

e) $P(A_1 \cap A_2 \cap A_3) = P(A_3 | A_2 \cap A_1) \cdot P(A_2 | A_1) P(A_1)$

RHS $= P(A_3 | A_2 \cap A_1) \cdot P(A_2 | A_1) \cdot P(A_1)$

$= \frac{P(A_3 \cap A_2 \cap A_1)}{\cancel{P(A_2 \cap A_1)}} \cdot \frac{\cancel{P(A_2 \cap A_1)}}{\cancel{P(A_1)}} \cdot \cancel{P(A_1)}$

$=$ LHS

True

Discrete & Continuos Distribution

Multivariate Gaussian

$$\frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2} - (x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

2. Bernoulli : $p^x (1-p)^{1-x}$
3. Uniform $\frac{1}{b-a}$ when $a \leq x \leq b$
   0 otherwise
4. Binominal $\binom{n}{x} p^x (1-p)^{n-x}$

Mean, Variance and entropy

@ $X \to$ random variable with expection value $< \infty$

Variance $\to V(x) = \mathbb{E}((X - \mathbb{E}X)^2)$

$\Rightarrow Var(X) = \mathbb{E}(X^2 + (\mathbb{E}X)^2 - 2X(\mathbb{E}X))$

$= \mathbb{E}(X^2) + \mathbb{E}(\mathbb{E}X)^2 - 2\mathbb{E}(X(\mathbb{E}X))$

$= \mathbb{E}(X^2) + (\mathbb{E}X)^2 - 2\,\mathbb{E}X\,\mathbb{E}(\mathbb{E}X)$

$= \mathbb{E}(X^2) + (\mathbb{E}X)^2 - 2(\mathbb{E}X)^2$

$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$

b) For Bernolli (p) random variable

$Pr(x=1) = p \quad Pr(x=0) = 1-p$

Mean $= Pr(x=1).1 + Pr(x=0).0$

$= pr(x=1)$

mean$(E(x)) = p$

Varianc$(x) = E(x^2) - (E(x))^2$

$E(x^2) = Pr(x=1)1^2 + Pr(x=0).0^2$

$= Pr(x=1)$

$= p$

$$Var(x) = p - p^2 = p(1-p)$$

$$Entropy(X) = -\sum_{i \in \{0,1\}} P(x_i) \log_2 P(x_i)$$

$$= -[Pr(x=0)\log_2 Pr(x=0) + Pr(x=1) \log_2 Pr(x=1)]$$

$$Entropy(X) = -(1-p) \cdot \log_2(1-p) - p \cdot \log_2 p$$

Law of Large Number and Central limit Theorem

ⓐ By law of large number $Pr(x=3) = \frac{1}{6}$

Then time 3 apeears is $6000 \times \frac{1}{6} = 1000$

b) Accoriding to central limit theorem

$$\sqrt{n}\,(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

for the fair coin $p = 1/2$

so $\mu = 1/2$ and $\sigma^2 = p(1-p) = 1/4$

Linear Algebra

Geometry

@ To show $\vec{\omega} \perp (\omega^T \vec{x} + b = 0)$

we have to show that for two poin x_1, x_2

on the line $\vec{\omega}^T (\vec{x}_1 - \vec{x}_2) = 0$

the points follows line equation, also

$$\vec{\omega}^T x_1 + b = 0$$

$$\omega^T x_2 + b = 0$$

Subtracting them we get

$$\omega^T (x_1 - x_2) = 0$$

$\Rightarrow$ inner product of $\omega^T (x_1 - x_2)$ is zero

$\Rightarrow \omega^T$ is $\perp$ to $x_1 - x_2$

$\rightarrow$ and as x_1 & x_2 are on the line $x_1 - x_2$ is $\parallel$ to the line $\omega^T x + b = 0$

$\Rightarrow \omega^T$ is $\perp$ to $\omega^T x + b$

b) $\vec{w}^T\vec{x} + b = 0$

If x is multi dimentional. then

$w_1 x_1 + w_2 x_2 + \ldots\ldots w_n x_n + b = 0$

where n is the dimension of x

Now distace from oridiin

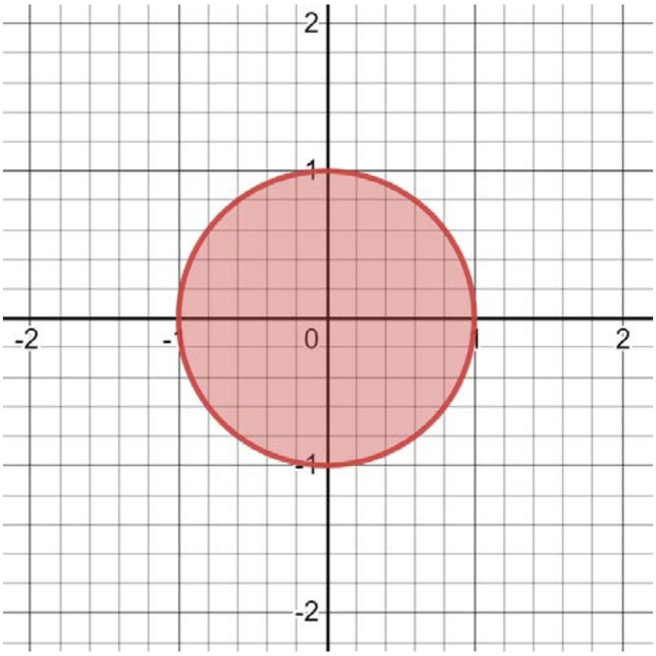
$$d = \frac{|w_1 \cdot 0 + w_2 \cdot 0 + \ldots + w_n \cdot 0 + b|}{\sqrt{w_1^2 + w_2^2 + \ldots + w_n^2}}$$

$$d = \frac{|b|}{||w||}$$

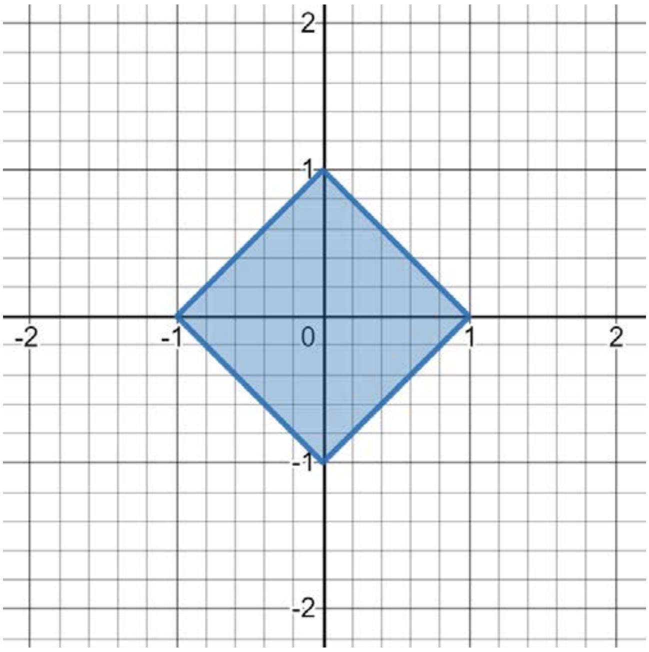
for $b > 0$ $d = \frac{b}{||w||}$

Vector norms

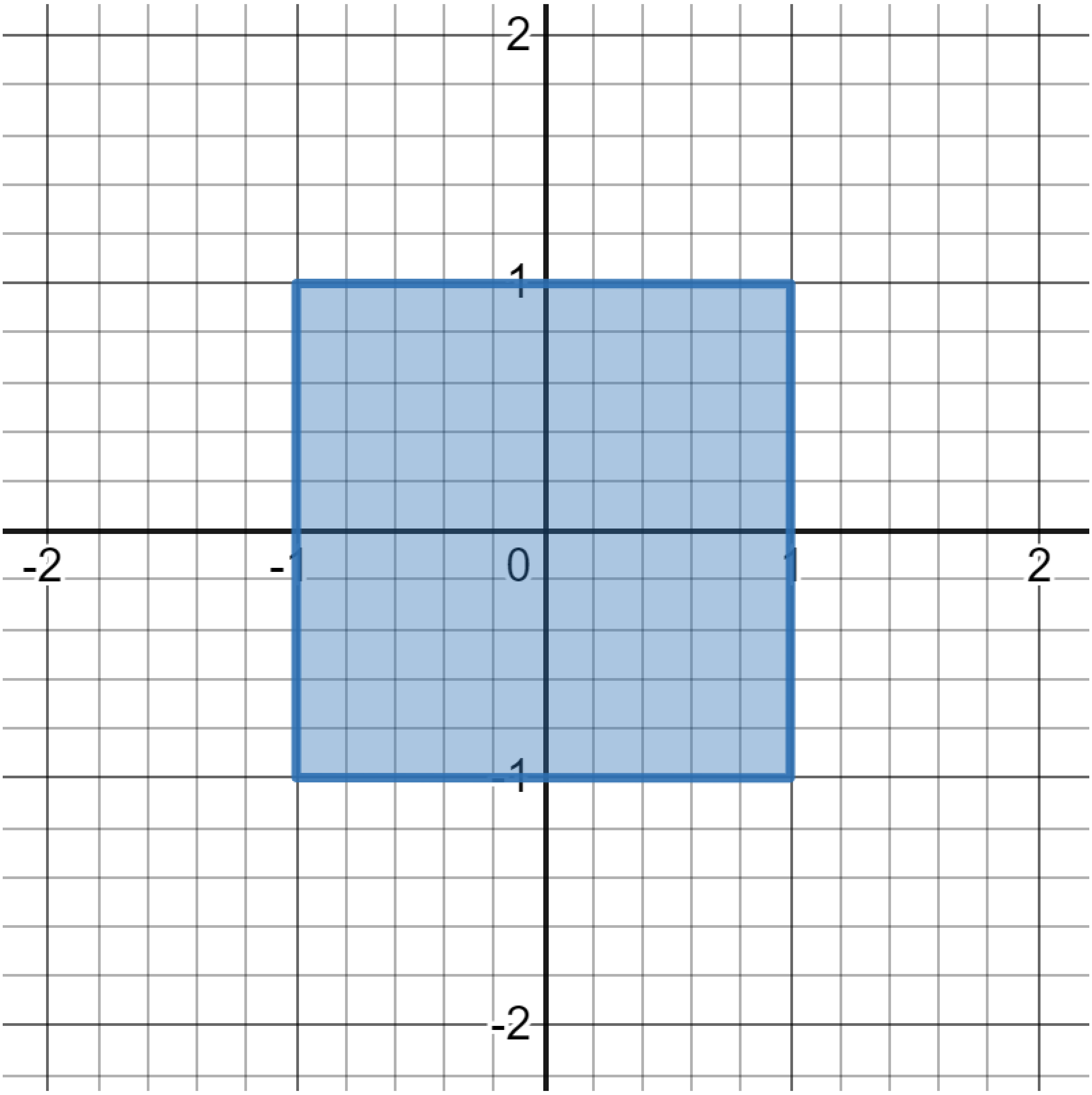
(a) $||x||_2 \leq 1$ $\quad$ $Here\ \sqrt{x^2 + y^2} \leq 1$



(b) $||X||_1 \leq 1$ $\quad$ $Here\ |x| + |y| \leq 1$



(b) $||X||\_\{\infty\} \leq 1$ $\quad$ $Here \max(|x|, |y|) \leq 1$



Homework_0

September 10, 2021

0.1 Programming Skills

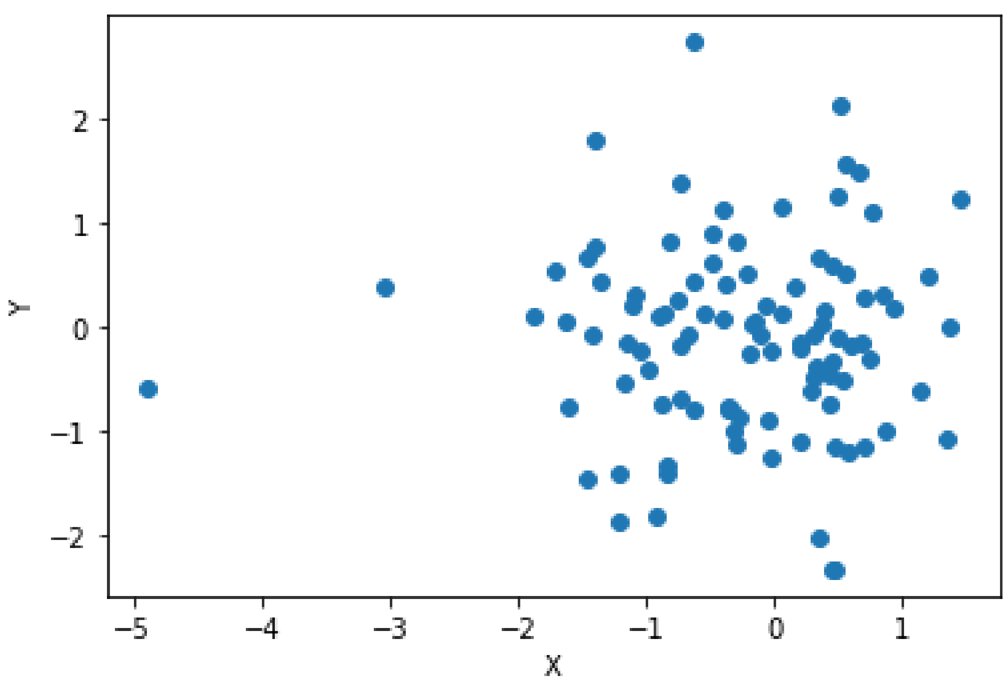
0.1.1 Sampling from a distribution. Use the Python libraries numpy and matplotlib

Draw 100 samples x = [x1 x2] from a 2-dimensional Gaussian distribution with mean [0,0] and identity covariance matrix. Plot them on a scatter plot (x1 vs. x2).

[34]:
```
import numpy as np
import matplotlib.pyplot as plt
```

[35]:
```
mean=np.array([0,0])
cov=np.array([[1,0],[0,1]])
x, y = np.random.multivariate_normal(mean, cov, 100).T
plt.scatter(x, y)
plt.xlabel('X')
plt.ylabel('Y')
```

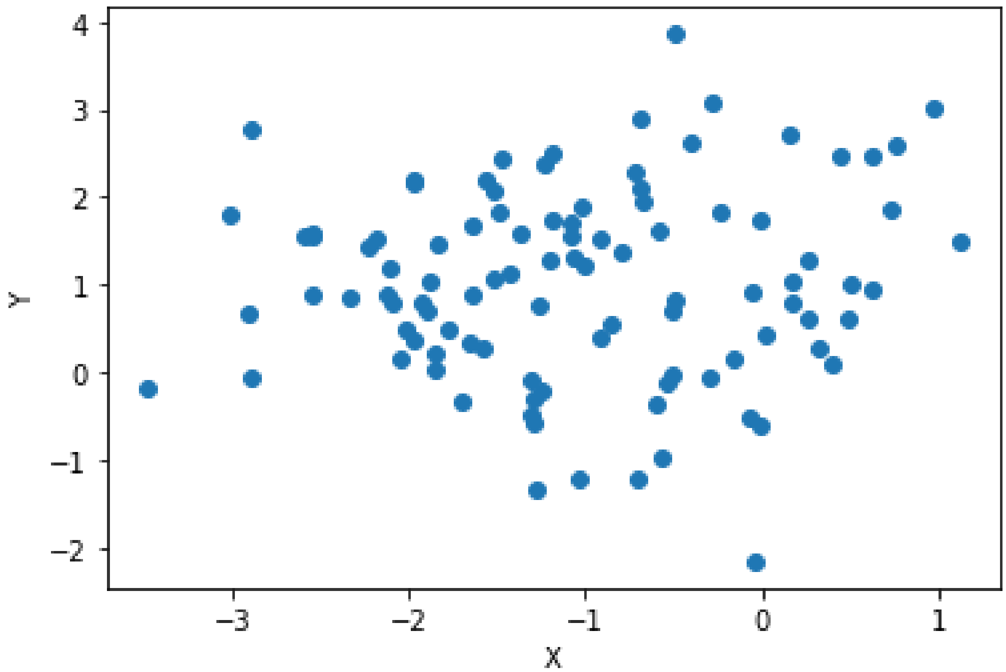
[35]: `Text(0, 0.5, 'Y')`



How does the scatter plot change if the mean is [-1, 1] ?

[36]:
```
mean=np.array([-1,1])
x, y = np.random.multivariate_normal(mean, cov, 100).T
plt.scatter(x, y)
plt.xlabel('X')
plt.ylabel('Y')
# Data seams more clustered and shift to -1 in y axis and +1 in the x axis
```

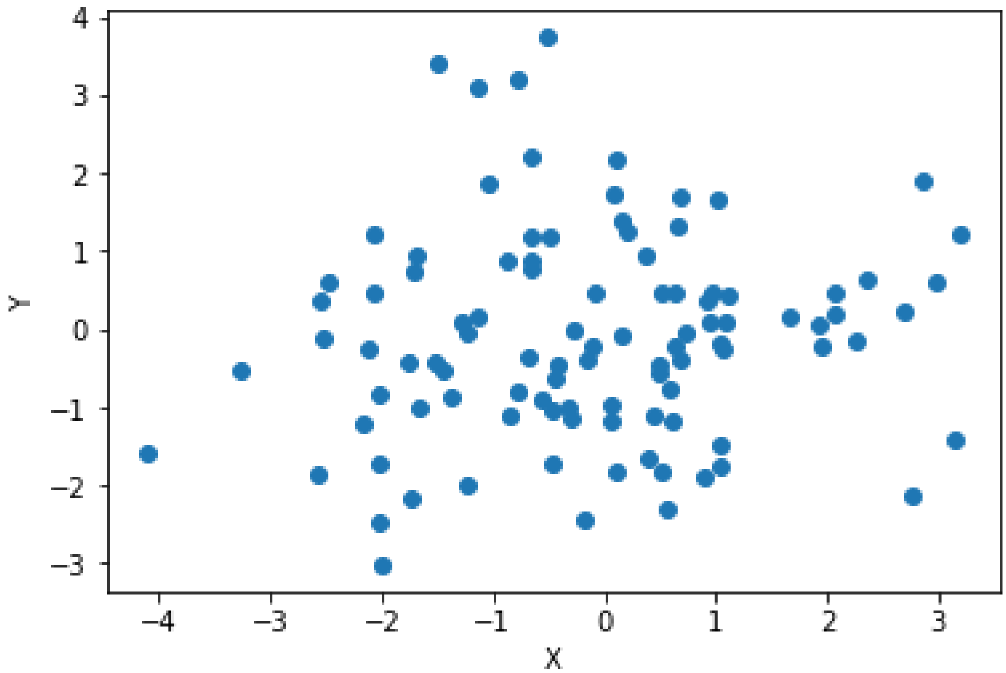
[36]: Text(0, 0.5, 'Y')



How does the scatter plot change if you double the variance of each component (x1 & x2)?

[37]:
```
mean=np.array([0,0])
cov=np.array([[2,0],[0,2]])
x, y = np.random.multivariate_normal(mean, cov, 100).T
plt.scatter(x, y)
plt.xlabel('X')
plt.ylabel('Y')
# The scatter is more uniform
```

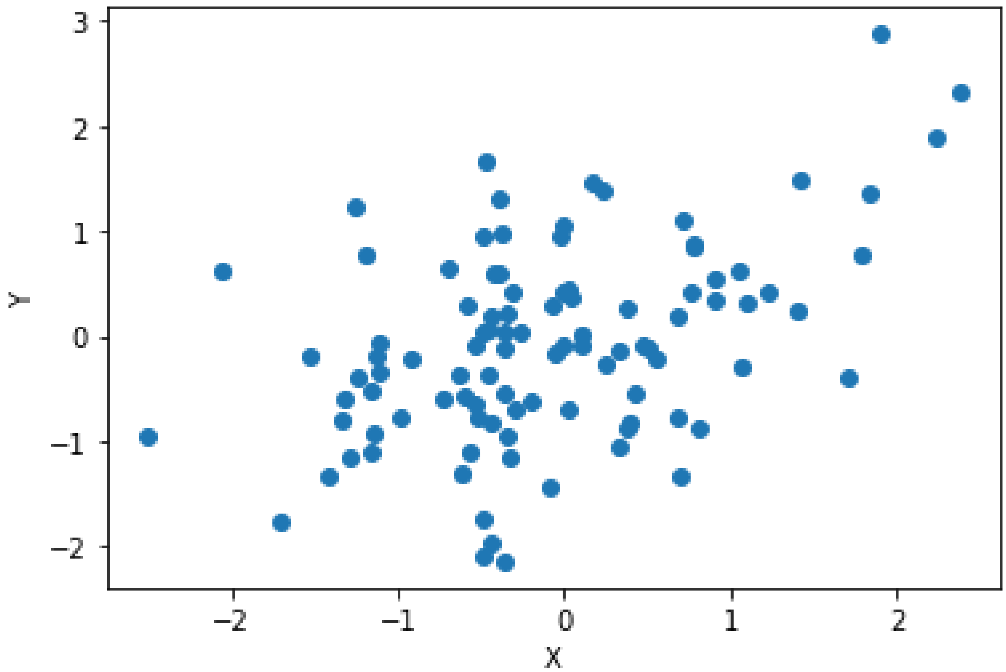
[37]: Text(0, 0.5, 'Y')



How does the scatter plot change if the covariance matrix is changed to the following? [[1,0.5], [0.5,1]]

[38]:
```
mean=np.array([0,0])
cov=np.array([[1,0.5],[0.5,1]])
x, y = np.random.multivariate_normal(mean, cov, 100).T
plt.scatter(x, y)
plt.xlabel('X')
plt.ylabel('Y')
# more clusterd data in 1st and 3rd quadrat
```

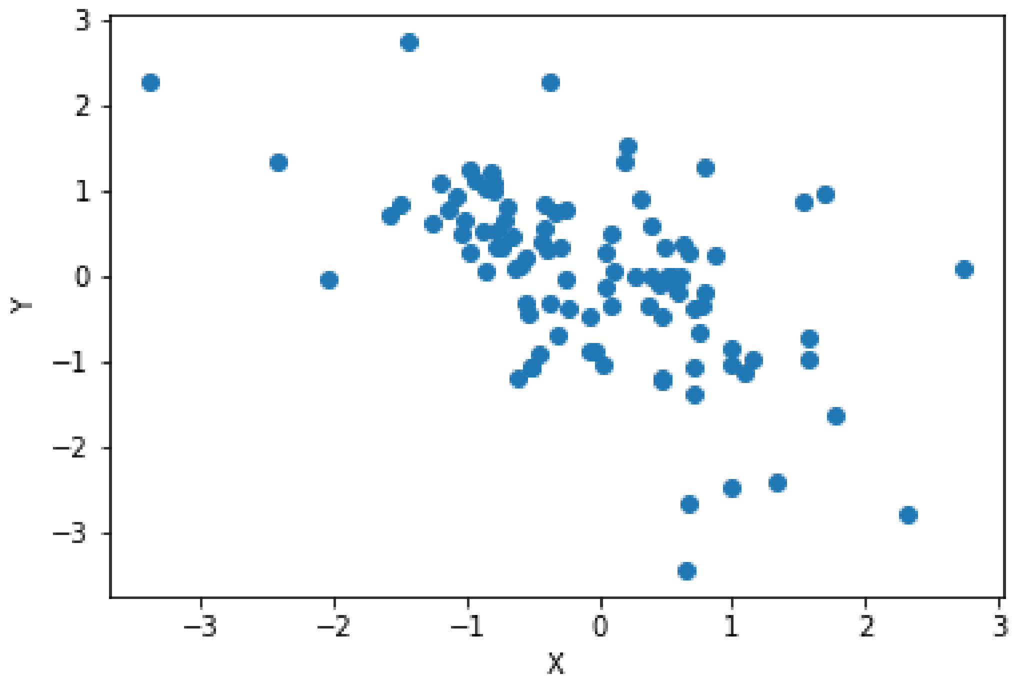
[38]:
```
Text(0, 0.5, 'Y')
```



How does the scatter plot change if the covariance matrix is changed to the following? [[1,-0.5],[-0.5,1]]

[39]:
```
mean=np.array([0,0])
cov=np.array([[1,-0.5],[-0.5,1]])
x, y = np.random.multivariate_normal(mean, cov, 100).T
plt.scatter(x, y)
plt.xlabel('X')
plt.ylabel('Y')
# more clusterd data in 2nd and 4th quadrant
```

[39]:
```
Text(0, 0.5, 'Y')
```



Taken help from Ajay sahoo for the math parts of the assignment

```
[ ]:
```