Name:- Prateek Mahajan

EE511 Advanced Intro to ML

Homework 1:-

Problem 1:-

a) I have read and understood the general instructions at the top of HW1 and I formally declare that all work I turn in for everything in this course will not contain or involve any cheating at all.

Problem 2:-

a) P/T $\frac{\partial x^T a}{\partial x} = a$

Take LHS:-

$x^T \rightarrow 1 \times m$ row vector

$a \rightarrow m \times 1$ column vector

Let $x^T = [x_1 \; x_2 \; x_3 \; \ldots \; x_m]$

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

$$\therefore LHS = \frac{\partial x^T a}{\partial x} = \frac{\partial}{\partial x}\left([x_1 \; x_2 \ldots x_m] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \right)$$

~~Consider~~ $\cancel{x^T a}$ ~~&~~ $\cancel{\frac{\partial}{\partial x}(x^T a)}$

~~Partial derivatives of x wrt elements of x can be written as :-~~

$$\cancel{\frac{\partial(x^T a)}{\partial x_i} = \frac{\partial(x^T a)}{\partial}}$$

Consider $x^T a$.

$x^T a$ outputs a matrix of size 1×1

$$\Rightarrow x^T a = \sum_{j=1}^{m} x_{1,j} \cdot a_{j,1}$$

$$\therefore \frac{\partial(x^T a)}{\partial x} = \frac{\partial\left(\sum_{j=1}^{m} x_{1,j} \times a_{j,1}\right)}{\partial x}$$

Since this is a derivative wrt a vector/matrix of size $m \times 1$:-

$$\text{LHS} = \frac{\partial(x^T a)}{\partial x} = \text{Jacobian}\left(\sum_{j=1}^{m} x_{1,j} \times a_{j,1}\right)_x$$

$$= \begin{bmatrix} \frac{\partial(x^T a)}{\partial x_1} \\ \frac{\partial}{\partial x_2}(x^T a) \\ \vdots \\ \frac{\partial}{\partial x_m}(x^T a) \end{bmatrix}$$

$$= \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{bmatrix} = a = \text{RHS}$$

Hence LHS = RHS $\Rightarrow$ this is proved.

b) P/T $\frac{\partial(a^T X b)}{\partial X} = ab^T$

$a^T \rightarrow 1 \times m$ row vector

$b \rightarrow m \times 1$ column vector

$X \rightarrow m \times m$ matrix.

$\therefore$ consider $a^T X b$.

$$a^T X b = [a_1 \; a_2 \; \ldots \; a_m] \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1m} \\ x_{21} & & & \vdots \\ \vdots & & & \vdots \\ x_{m1} & x_{m2} & \ldots & x_{mm} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}$$

$$= \left[\sum_{i=1}^{m} a_i \cdot x_{i,1} \quad \sum_{i=1}^{m} a_i x_{i,2} \quad \ldots \quad \sum_{i=1}^{m} a_i x_{i,m} \right] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$$= \sum_{j=1}^{m} \left(\Sigma a_i x_{i,j} \right) b_j$$

$$\therefore LHS = \frac{\partial \left(\sum_{j=1}^{m} (\Sigma a_i x_{i,j}) b_j \right)}{\partial X}$$

Since X is a matrix this is a jacobian

$$\therefore LHS = \begin{bmatrix} \frac{\partial(a^T X b)}{\partial x_{11}} & \frac{\partial(a^T X b)}{\partial x_{12}} & \frac{\partial(a^T X b)}{\partial x_{13}} & \ldots & \frac{\partial(a^T X b)}{\partial x_{1m}} \\ \frac{\partial(a^T X b)}{\partial x_{21}} & & & & \vdots \\ \vdots & & & & \vdots \\ \frac{\partial(a^T X b)}{\partial x_{m1}} & \ldots & & & \frac{\partial(a^T X b)}{\partial x_{mm}} \end{bmatrix}$$

$$= \begin{bmatrix} a_1 b_1 & a_1 b_2 & a_1 b_3 & \ldots & a_1 b_m \\ a_2 b_1 & a_2 b_2 & & & \vdots \\ a_3 b_1 & & & & \vdots \\ \vdots & & & & \vdots \\ a_m b_1 & & & & a_m b_m \end{bmatrix}$$

$$= a b^T = RHS$$

$\therefore LHS = RHS \Rightarrow$ 2b) is proved.

Problem 3:-

a) $W \times \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$

$\Rightarrow W \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 3 & 2 \end{bmatrix}$

$\Rightarrow W = \begin{bmatrix} 2 & 0 \\ 3 & 2 \end{bmatrix}$

a) $a_1 . W = b_1 \Rightarrow [1 \quad 0] . W = [2 \quad 0]$

$a_2 . W = b_2 \Rightarrow [0 \quad 1] . W = [3 \quad 2]$

Let W be $\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$

a) ~~let a~~ $W a_1 = b_1$; $W a_2 = b_2$

$\Rightarrow W . \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$; $W . \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

Let $W = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$

$\therefore w_{11} = 2$; $w_{21} = 0$; $w_{12} = 3$; $w_{22} = 2$

(multiplying the above eqns)

$\therefore W = \begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix}$

b) For V $\tan\alpha = 2$

To rotate ~~a point~~ around origin counterclockwise,)

R = rotation matrix $= \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$

$\Rightarrow$ to rotate by α (s/t $\tan\alpha = 2$) in CC is equiv. to rotating α (s/t $\tan\alpha = -2$) in the clockwise direction

$\tan\alpha = -2 \Rightarrow \sin\alpha = -\frac{2}{\sqrt{5}}$, $\cos\alpha = \frac{1}{\sqrt{5}}$

$\therefore$ rotation matrix, $V = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}$

Σ scales the x-coordinate of a point by 4

$\Rightarrow \Sigma \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4x \\ y \end{bmatrix}$ $\left(\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$

$\Rightarrow \Sigma_{11} x + \Sigma_{12} y = 4x$

$\Sigma_{21} x + \Sigma_{22} y = y$

$\Rightarrow \Sigma_{11} = 4, \Sigma_{22} = 1, \Sigma_{12} = \Sigma_{21} = 0$

$\therefore \Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$

For U's case, it is a rotation matrix of the form, s/t it rotates by β in CC s/t $\tan\beta = 1/2$.

$\therefore \sin\beta = \frac{1}{\sqrt{5}}$, $\cos\beta = \frac{2}{\sqrt{5}}$

~~$\therefore U = \begin{bmatrix} 3/\sqrt{10} & -1/\sqrt{10} \\ 1/\sqrt{10} & 3/\sqrt{10} \end{bmatrix}$~~

~~$\therefore U \Sigma V = \begin{bmatrix} 3/\sqrt{10} & -1/\sqrt{10} \\ 1/\sqrt{10} & 3/\sqrt{10} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}$~~

~~$= \begin{bmatrix} 12/\sqrt{10} & -1/\sqrt{10} \\ 4/\sqrt{10} & 3/\sqrt{10} \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}$~~

~~$= \begin{bmatrix} 14/(2.5) & 23/5\sqrt{2} \\ -2/5\sqrt{2} & 11/5\sqrt{2} \end{bmatrix}$~~

$\therefore U = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$

$\therefore U.\Sigma.V = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ +1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & +2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}$

$$= \begin{bmatrix} 8/\sqrt{5} & -1/\sqrt{5} \\ +4/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & +2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix}$$

this is the same as W in qn. 2a)

c) $W = \begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix}$

$$\therefore W^T W = \begin{bmatrix} 2 & 0 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 6 & 13 \end{bmatrix}$$

let $Z = W^T W$

To find eigenvalues $AZ = \lambda Z$

$\Rightarrow$ first set $|A - \lambda I| = 0$

$$\Rightarrow \begin{vmatrix} 4-\lambda & 6 \\ 6 & 13-\lambda \end{vmatrix} = 0 \Rightarrow 52 - 17\lambda + \lambda^2 - 36 = 0$$

$$\Rightarrow \lambda^2 - 17\lambda + 16 = 0$$

$$\Rightarrow (\lambda - 16)(\lambda - 1) = 0$$

$$\Rightarrow \lambda = 1, 16$$

$\therefore$ to find eigenvectors :-

do $(A - \lambda I).v = 0$

For $\lambda = 1$:

$$\begin{bmatrix} 3 & 6 \\ 6 & 12 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$v_1 = -2v_2 \Rightarrow$ eigenvector $= \left(\frac{-2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right)$. (normalised form)

For $\lambda = 16$:-

$$\begin{bmatrix} -12 & 6 \\ 6 & -3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$\Rightarrow v_2 = 2v_1$

$\therefore$ a possible eigenvector $= \left(\frac{1}{\sqrt{5}} \; \frac{2}{\sqrt{5}}\right)$ (normalised form)

A point on the unit circle can be written as $\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$

$\therefore W. \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} = \begin{bmatrix} 2\cos\theta + 3\sin\theta \\ 2\sin\theta \end{bmatrix}$

$\therefore$ transformed point $P = \begin{bmatrix} 2\cos\theta + 3\sin\theta \\ 2\sin\theta \end{bmatrix}$

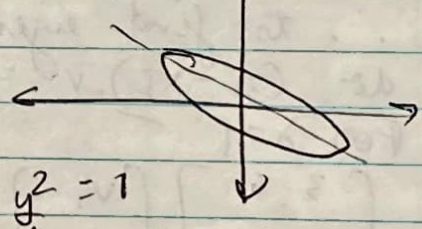
Based on the answer 3b); transforming with W is equiv. to rotating clockwise by $\tan^{-1}(2)$, scaling x-coordinate by 4 & rotating counter clockwise by $\tan^{-1}(1/2)$.

→ The first rotation has no impact, as rotating a circle doesn't change it.

→ On scaling the x-val. by 4, this would become an ellipse with major axis 8 & minor axis 2

→ The final rotation would give a rotated ellipse with same major & minor axes, & making an angle $\tan^{-1}(1/2)$ with x-axis

approx. final ellipse



The eq^n of the ellipse $\frac{(x-3y)^2}{4} + \frac{y^2}{4} = 1$

$\Rightarrow \frac{x^2 - 6xy + 10y^2}{4} = 1$

Length of semi-major axis (a) $= 8/2 = 4$

Length of semi-minor axis (b) $= 2/1 = 1$

Clearly, these are equal to $\sqrt{\lambda}$ ($4 = \sqrt{16}$; $1 = \sqrt{1}$)

square root of eigenvalues

d) $W = \begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix}$ $\therefore$ $\det(W) = \begin{vmatrix} 2 & 3 \\ 0 & 2 \end{vmatrix} = \underline{\underline{4}}$

The area of the transformed unit circle = area of ellipse $= \pi \times a \times b = \underline{\underline{4\pi}} \simeq \underline{\underline{12.56 \text{ unit}^2}}$

$\therefore$, clearly area of transformed figure = $\det(W)$ * area of unit circle

(since area of unit circle $= \pi$)

~~Clearly~~ in

The determinant can, therefore kind of be seen as "space" enclosed by various vectors or the matrix.

Since this is a 2×2 matrix, "space" corresponds to area.

$\therefore$ since area of transformed fig. = $\det(W)$ * area of unit circle we can say that this is equivalent to saying that $\det(AB) = \det(A) * \det(B)$, for 2 matrices A & B, which is why the given impression should be true

Problem 4:-

a) P(my win on machine 1) = 0.4

P(friend win on machine 1) = 0.3

P(my win on machine 2) $= \frac{210}{1040} = 0.202$

P(friend win on machine 2) $= \frac{14}{84} = \frac{1}{6} = 0.167$

On ~~both~~ either machine 1 & machine 2 individually, I am more likely to win

b) P(me winning in casino) $= \frac{250}{1140} = 0.219$

P(friend winning in casino) $= \frac{\cancel{44}\,11}{\cancel{18}\,46} = 0.239$

Based on the values taking all runs on machine 1 & 2 together, my friend is more likely to win.

c) As per 4a) I am more likely to win either machine individualy

As per 4b) my friend is more likely to win in the casino.

Therefore, despite my winning prob. being higher on both machines, my ~~total~~ winning probability is less than my friends.

This can be simply be attributed to the fact that I ~~spent~~ played far more rounds on the lower probability slot machine, and so, took a large hit on my final winning probability

~~P(winning in casino) = P(winning in machine 1) + P(winning in machine 2)~~

$$P(\text{winning in casino}) = \frac{\overbrace{P(\text{winning in M1}) * (\text{rounds in M1})}^{T_A} + \underbrace{P(\text{winning in M2}) * (\text{rounds in M2})}_{T_B}}{\text{Total Rounds in } M_1 + \text{Rounds in M2}}$$

In my case "rounds in M2" was high, so T_B was high & brought my net winning probability down.
However, in my friend's case "rounds in M1" & "rounds in M2" were comparable, so the impact of a lower P(winning in M2) ~~didn't~~ wasn't as significant as for me. This lead to the results ~~explained~~ noticed above.

d) Assuming that the probabilities remain the same for me & my friend on both machines a) & b) give the same conclusion when my friend & I run a similar no. of rounds on both machines. Variations in the ratio of no. of rounds run on both machines may/may not affect the conclusions, depending on how significant the difference in the ratio is b/w me & my friend.

Problem 5 :-

a) Mean of multivariate gaussian distribution = [0.01048 , 0.06125]
Covariance matrix of generated distribution =

$$\begin{bmatrix} 1.0357 & 1.5698 \\ 1.5698 & 5.0804 \end{bmatrix}$$

∴ covariance of variables of the multivariate distribution = 1.5698

b) please refer to the jupyter notebook

c) Yes, from the histograms, it is evident that both the x & y coordinates of X follow a gaussian distribution. I believe this is the case since there is a clear peak in the distribution ~~after~~ which drops as you go either side of it. (which is ~~the~~ characteristic of gaussian RVs)

$\therefore$ mean of X coordinate $= 0.01048$

mean of Y coordinate $= 0.06125$

variance of X coordinate $= 1.0357$

variance of Y coordinate $= 5.0804$

d) The difference in the scatter plots is that the plot generated by a multivariate gaussian distribution is tilted in comparison to the single variable gaussian distribution

This difference is caused by the fact that multivariate gaussian distributions ~~[illegible]~~ are ~~[illegible]~~ correlated to one another, while the individual gaussian distributions are not.

e) refer to jupyter notebook

f) From the histogram, clearly, the x-coordinates of the projected points seem like they have been sampled from a gaussian distribution.

$\therefore$ mean of X coordinates $= -0.2806$

variance of X coordinates $= 0.5618$

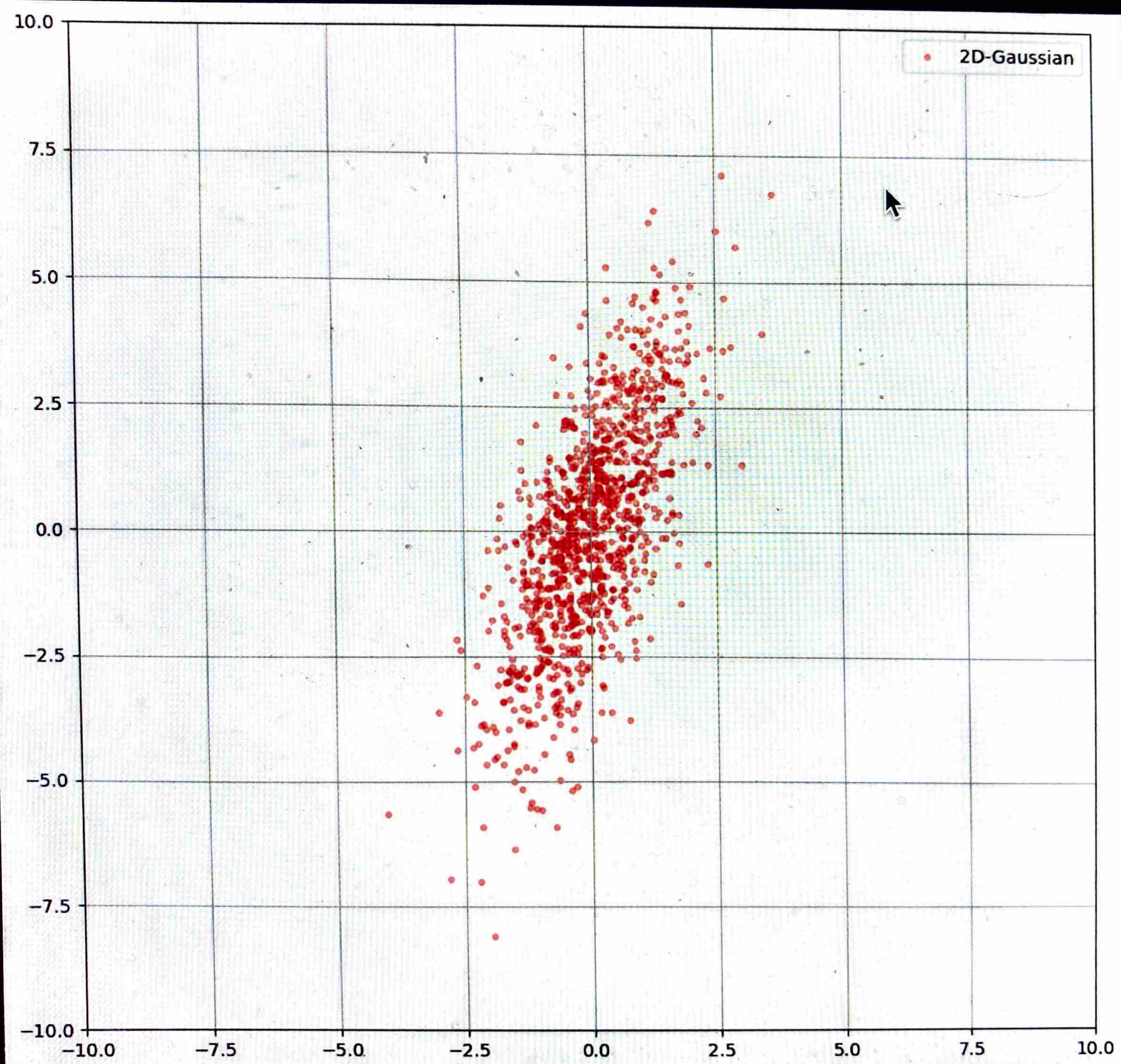
6 Problem 6:-

a) ~~TPR~~ Classifier A :-
TPR = 0.7273, FPR = 0.6667
Classifier B :-
TPR = 0.9091, FPR = 0.1111

b) refer to jupyter notebook

c) Classifier A :-
AUC = 0.7273
Classifier B :-
AUC = 0.9596

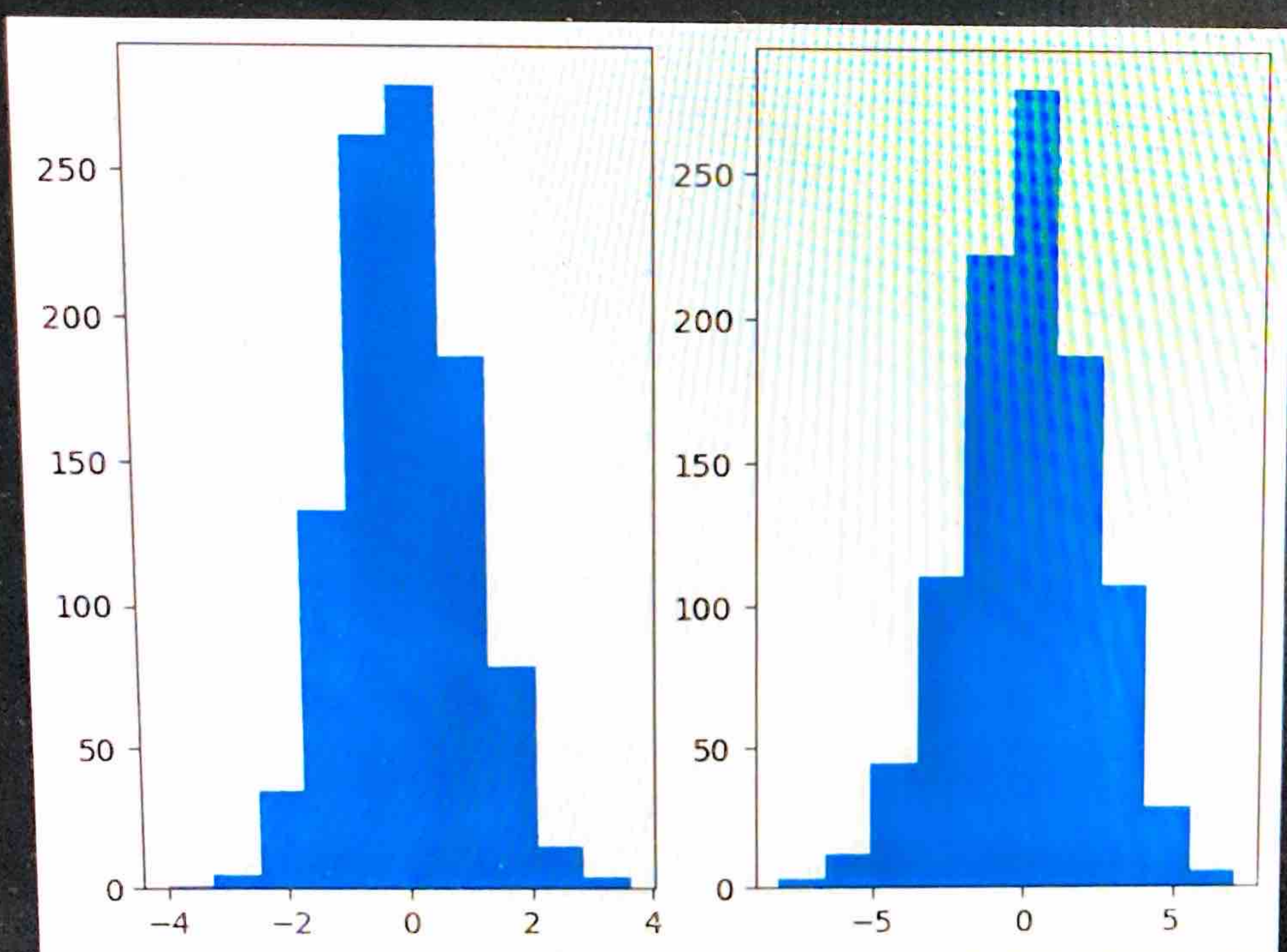
d) Classifier B clearly has a higher TPR, lower FPR & higher AUC, which are the desired characteristics of a classifier.
∴ B performs better on the given data


2D-Gaussian
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0

```
Mean of X:  [0.0104783  0.06124546]
Covariance :  [[1.03574467 1.56979754]
 [1.56979754 5.08040865]]
```

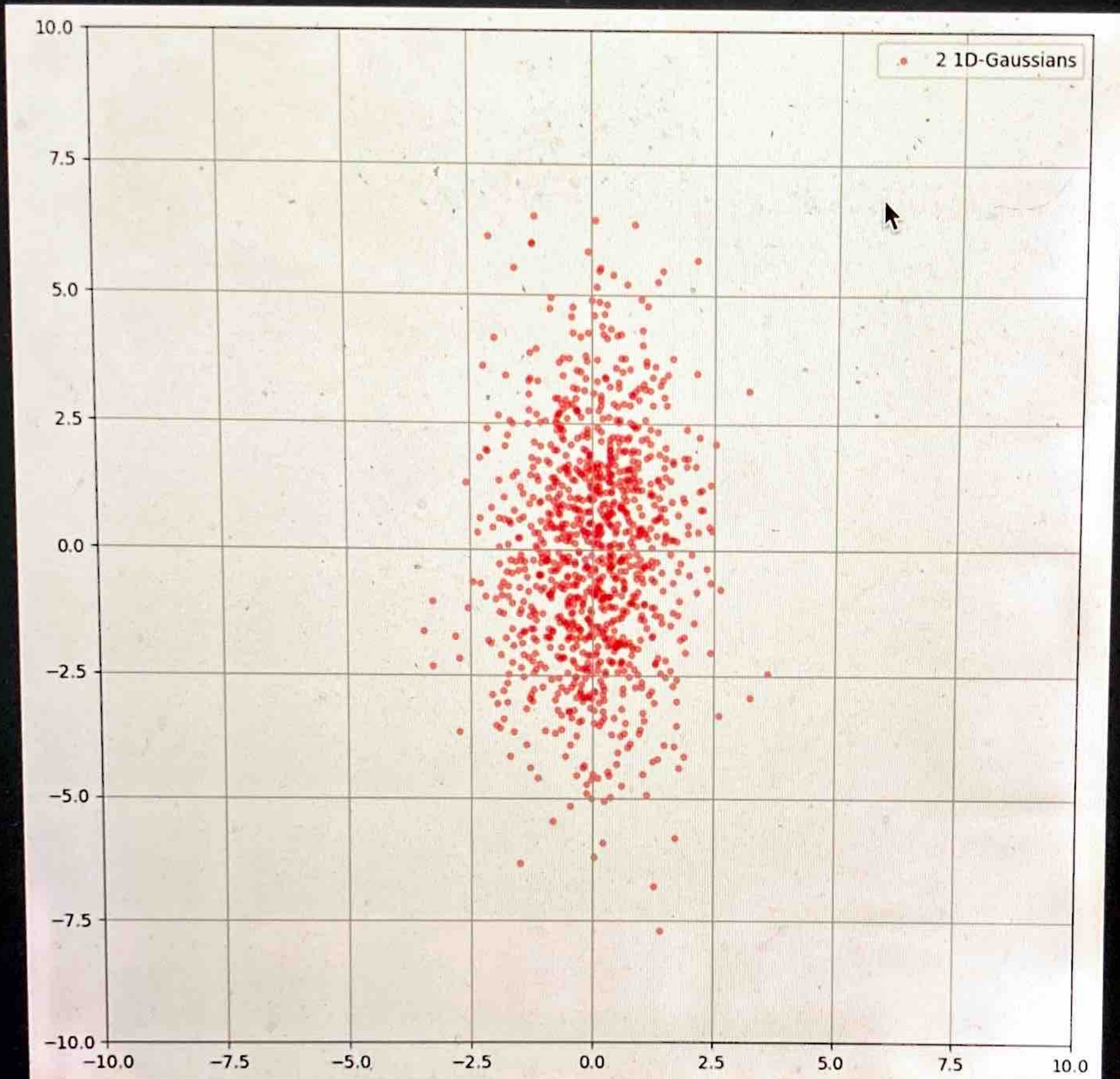
```
# TODO: Plot the histogram for the x-coordinates of X
# and y-coordinates of X respectively.
# You can use the plt.hist() function
plt.subplot(1,2,1)
plt.hist(X[:,0])
plt.subplot(1,2,2)
plt.hist(X[:,1])
```

```
(array([  3.,  12.,  45., 110., 222., 279., 187., 107.,  29.,   6.]),
 array([-8.09073464, -6.57100486, -5.05127508, -3.53154531, -2.01181553,
        -0.49208575,  1.02764403,  2.54737381,  4.06710359,  5.58683337,
         7.10656314]),
 <BarContainer object of 10 artists>)
```



[33]

```
Mean of X Coordinate:  0.01047830014298006
Mean of Y Coordinate:  0.061245463092547824
Covariance :  [[1.03574467 1.56979754]
 [1.56979754 5.08040865]]
/var/folders/c9/3q_3q36n4k9_t4gnwrhyd7s80000gn/T/ipykernel_92747/2508210211.py:28: UserWarning: FigureCanvasAgg is non-interactive, and thus cannot be shown
  fig.show()
```

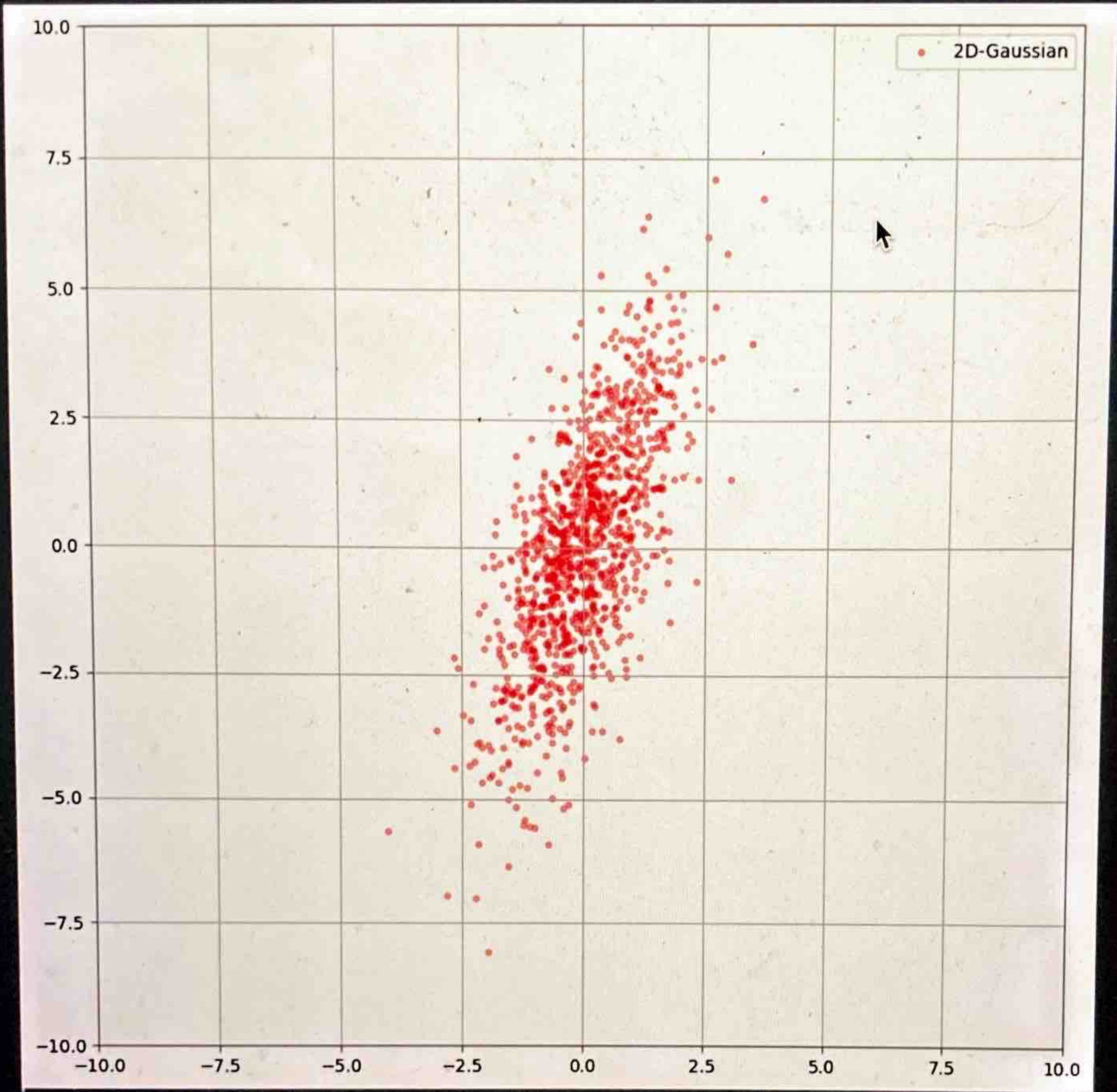



```
# Back to the original X
fig, ax = plt.subplots(figsize=(10, 10))
# c='r', dot color is red
```

```
fig.show()
```

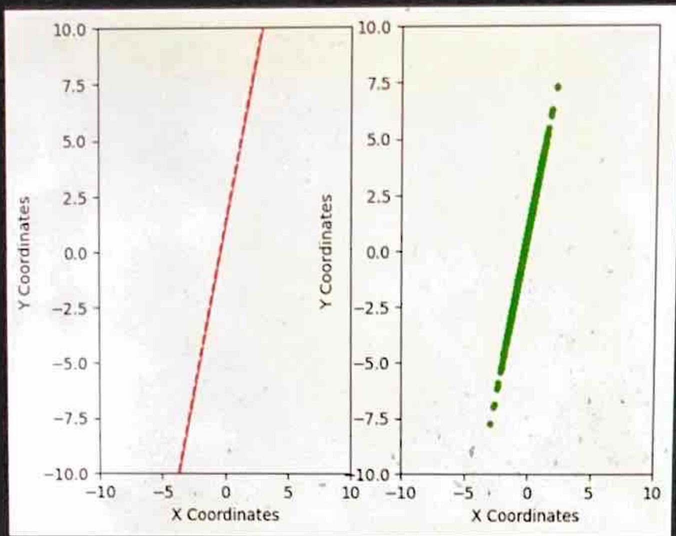
[25]

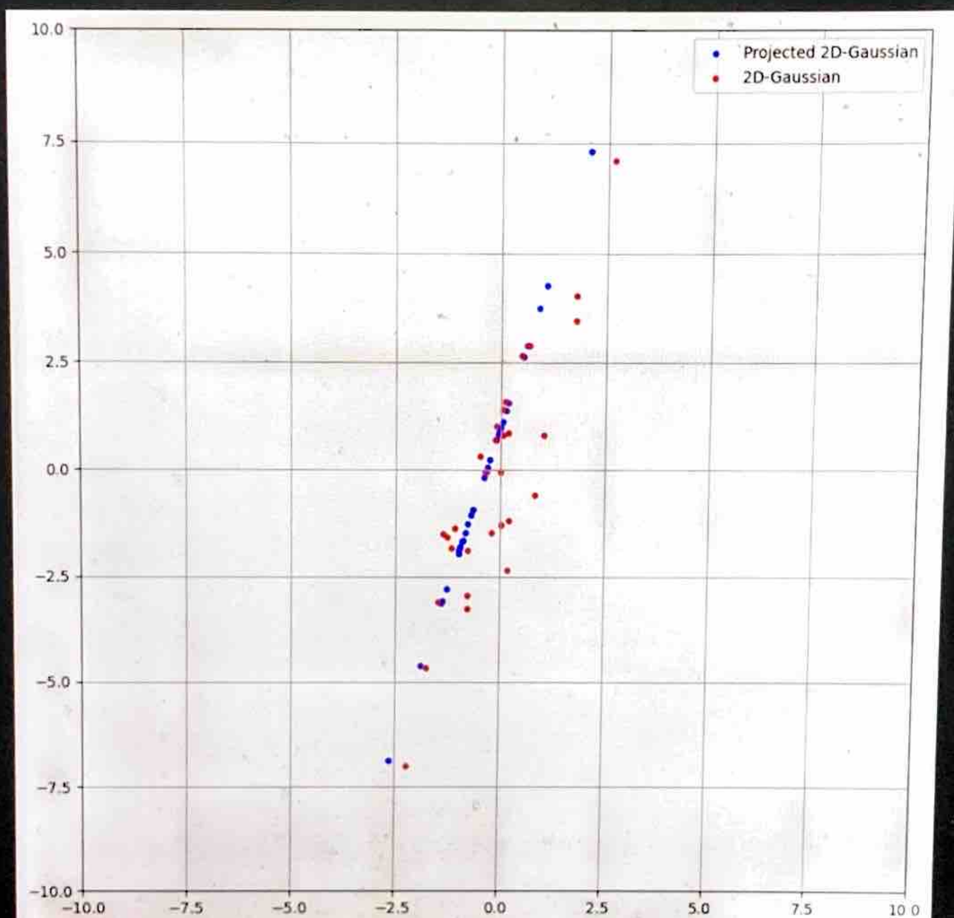
```
/var/folders/c9/3q_3q36n4k9_t4gnwrhyd7s80000gn/T/ipykernel_92747/3864500534.py:16: UserWarning: FigureCanvasAgg is non-interactive, and thus cannot
  fig.show()
```



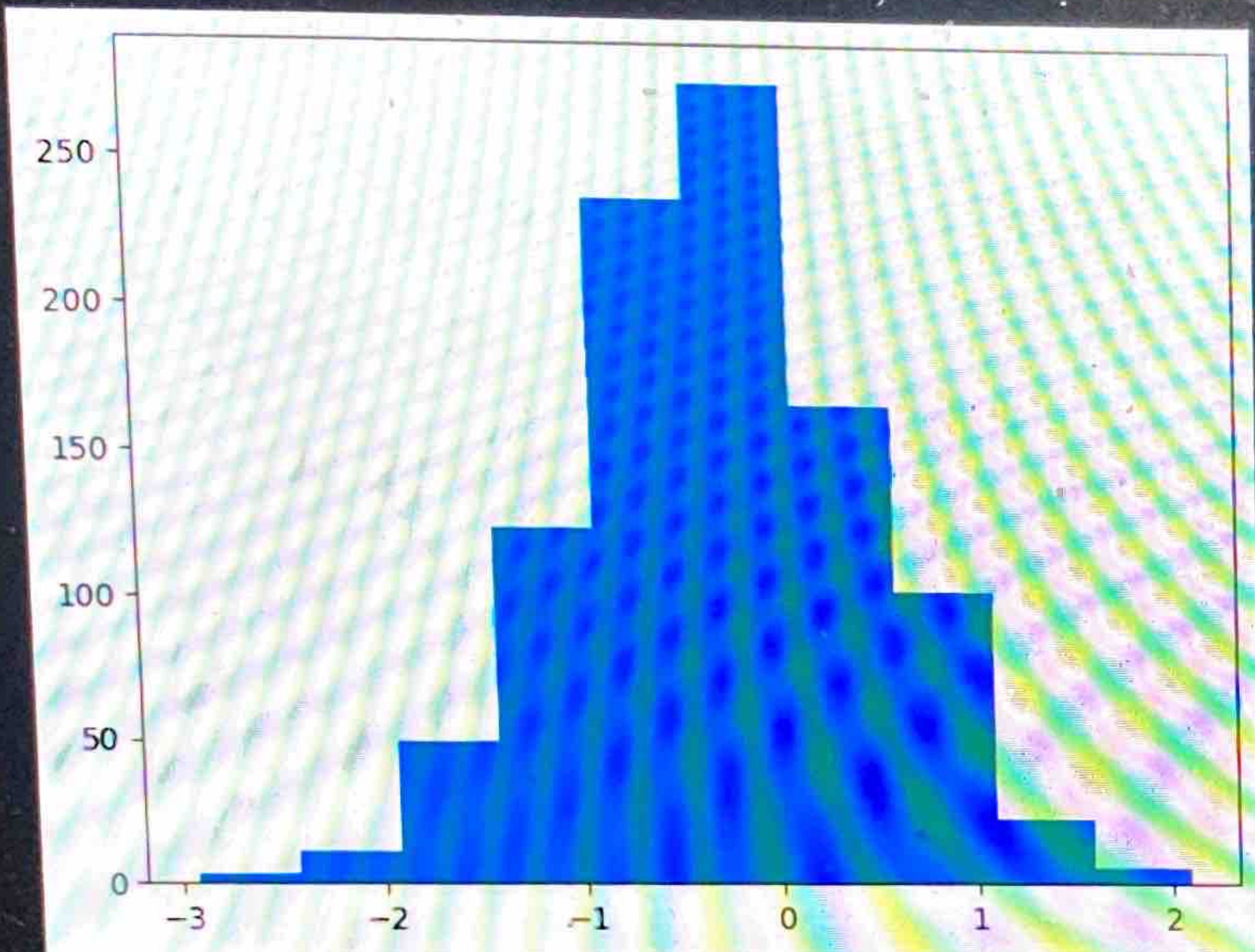

```
# TODO: Project X onto line y=3x + 1
```

```
/var/folders/c9/3q_3q36n4k9_t4gnwrhyd7s80000gn/T/ipykernel_92747/4099930313.py:36: UserWarning: FigureCanvasAgg is non-interactive, and thus cannot be shown
  fig.show()
```





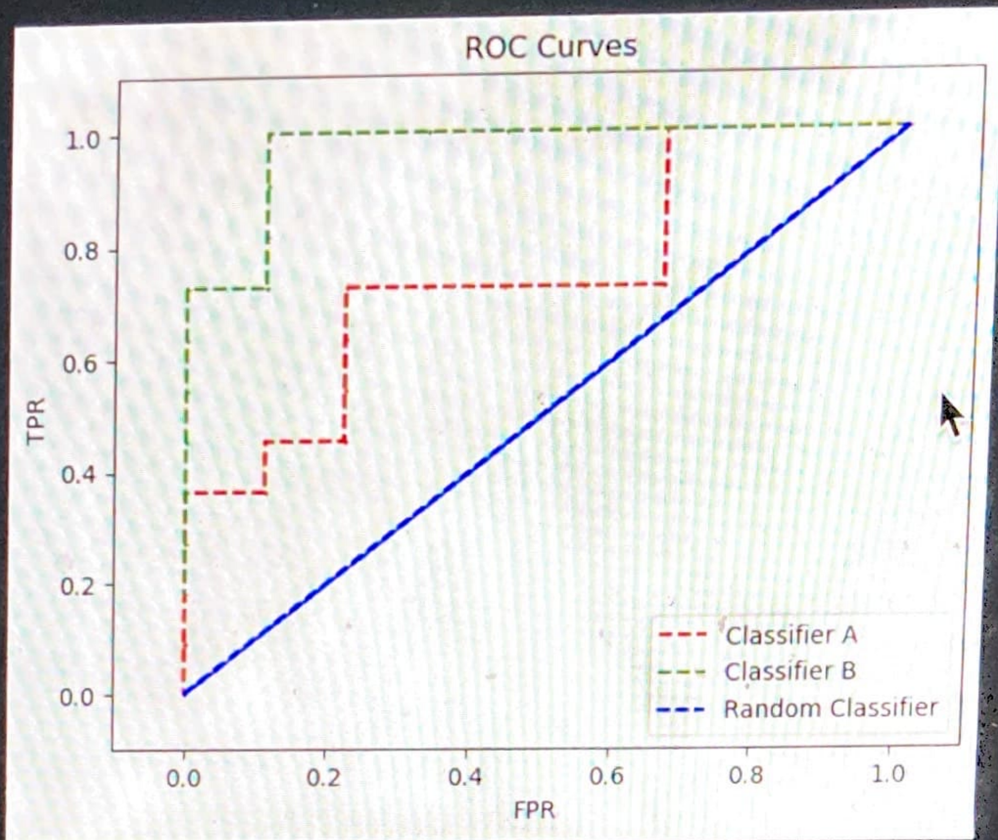
```
Mean of X Coordinate :  -0.2805785310579377
Variance of X Coordinate :  0.5617820770985318
```



```
Classifier A: TPR = 0.7273, FPR = 0.6667
Classifier B: TPR = 0.9091, FPR = 0.1111
```

[67]

```
<matplotlib.legend.Legend at 0x16b42afd0>
```




```python
def compute_auc(tpr, fpr):
    # Your code here
    # can use np.trapz to calculate the area under the ROC curve
    indexes = np.argsort(fpr)

    sorted_fpr = fpr[indexes]
    sorted_tpr = tpr[indexes]

    area = np.trapz(sorted_tpr, sorted_fpr)
    return area
```

[68]

```python
# Compute AUC
auc_value_a = compute_auc(np.array(tpr_a_list), np.array(fpr_a_list))
auc_value_b = compute_auc(np.array(tpr_b_list), np.array(fpr_b_list))
print("Classifier A: AUC value: {:.4f}".format(auc_value_a))
print("Classifier B: AUC value: {:.4f}".format(auc_value_b))
```

[69]

```
Classifier A: AUC value: 0.7273
Classifier B: AUC value: 0.9596
```

```python
from sklearn.metrics import roc_auc_score

auc_a = roc_auc_score(y_true, y_pred_a)
auc_b = roc_auc_score(y_true, y_pred_b)
print("Classifier A: AUC value (using scikit learn): {:.4f}".format(auc_a))
print("Classifier B: AUC value (using scikit learn): {:.4f}".format(auc_b))
```

[70]

```
Classifier A: AUC value (using scikit learn): 0.7475
Classifier B: AUC value (using scikit learn): 0.9697
```