Question 1 **(20 points)** Consider the following dataset used to train a binary decision tree to predict if the weather is good for playing outside. Use the dataset to answer the following questions. Show your work.

Outlook	Temperature	Windy	Play / Don't Play
sunny	85	false	Play
sunny	80	true	Don't Play
overcast	83	false	Play
rain	70	false	Play
rain	68	false	Don't Play
rain	65	true	Don't Play
overcast	64	true	Play
sunny	72	false	Don't Play
sunny	69	false	Play
rain	75	false	Play
sunny	75	true	Don't Play
overcast	72	true	Play
overcast	81	false	Play
rain	71	true	Don't Play

(a) (5 points) Using Gini impurity, determine the best splitting threshold for Temperature, out of the following values: [65, 70, 75, 80].

temp ≤ 65

T: 1 1   F: 7 5

0.4861

$LL = \frac{1}{2}\left(1-\left(\frac{1}{2}\right)^2-\left(\frac{1}{2}\right)^2\right) = 0.25$

$RL = \frac{1}{2}\left(1-\left(\frac{7}{12}\right)^2-\left(\frac{5}{12}\right)^2\right) = 0.243$

$Total = \frac{2}{14}(0.25) + \frac{12}{14}(0.243) = 0.244$

temp ≤ 70

T: 3 2   F: 5 4

$LL = \frac{1}{2}\left(1-\left(\frac{3}{5}\right)^2-\left(\frac{2}{5}\right)^2\right) = 0.24$

$RL = \frac{1}{2}\left(1-\left(\frac{5}{9}\right)^2-\left(\frac{4}{9}\right)^2\right) = 0.2469$

$Total = \frac{5}{14}(0.24) + \frac{9}{14}(0.2469) = 0.244$

Temp ≤ 75

T: 5 5   F: 3 1

$LL = \frac{1}{2}\left(1-\left(\frac{5}{10}\right)^2-\left(\frac{5}{10}\right)^2\right) = 0.25$

$RL = \frac{1}{2}\left(1-\left(\frac{3}{4}\right)^2-\left(\frac{1}{4}\right)^2\right) = 0.1875$

$total = \frac{10}{14}(0.25) + \frac{4}{14}(0.1875) = 0.2321$

temp ≤ 80

T: 5 6   F: 3 0

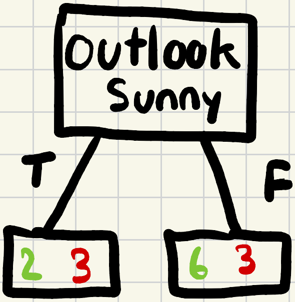
$LL = \frac{1}{2}\left(1-\left(\frac{5}{11}\right)^2-\left(\frac{6}{11}\right)^2\right) = 0.24795$

$RL = \frac{1}{2}\left(1-\left(\frac{3}{3}\right)^2-\left(\frac{0}{3}\right)^2\right) = 0$

$total = \frac{11}{14}(0.24795) + 0 = 0.1948$

The best Splitting treshold is 80

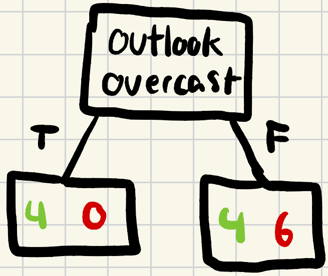
(b) (5 points) Analyze the impurity of the three features and determine the best feature for splitting.



$$LL = \frac{1}{2}\left(1-\left(\frac{2}{5}\right)^2-\left(\frac{3}{5}\right)^2 = 0.24$$

$$RL = \frac{1}{2}\left(1-\left(\frac{6}{9}\right)^2-\left(\frac{3}{9}\right)^2\right) = 0.22$$

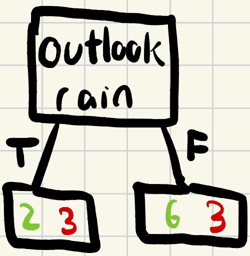
$$Tot = \frac{5}{14}(0.24)+\frac{9}{14}(0.22) = 0.227$$



$$LL = 0$$

$$RL = \frac{1}{2}\left(1-\left(\frac{4}{10}\right)^2-\left(\frac{6}{10}\right)^2\right) = 0.24$$

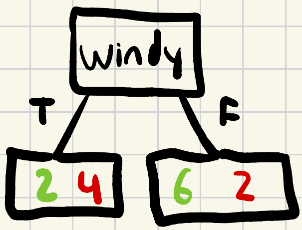
$$Tot = \frac{10}{14}(0.24) = 0.1714$$



$$LL = \frac{1}{2}\left(1-\left(\frac{2}{5}\right)^2-\left(\frac{3}{5}\right)^2 = 0.24$$

$$RL = \frac{1}{2}\left(1-\left(\frac{6}{9}\right)^2-\left(\frac{3}{9}\right)^2\right) = 0.22$$

$$Tot = \frac{5}{14}(0.24)+\frac{9}{14}(0.22) = 0.227$$



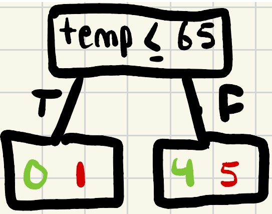
$$LL = \frac{1}{2}\left(1-\left(\frac{2}{6}\right)^2-\left(\frac{4}{6}\right)^2\right) = 0.22$$

$$RL = \frac{1}{2}\left(1-\left(\frac{6}{8}\right)^2-\left(\frac{2}{8}\right)^2\right) = 0.1875$$

$$Tot = \frac{6}{14}(0.22)+\frac{8}{14}(0.1875) = 0.2014$$

The best feature out of outlook, windy and temp is outlook/overcast w/ 0.1714 (lowest impurity)

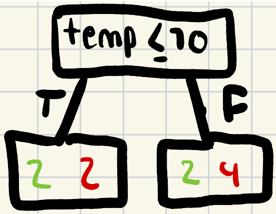
(c) (10 points) Finalize your decision tree. At each node, repeat the two previous steps to determine the best splitting feature and, if applicable, its threshold. Use a maximum depth of 4 (first split occurs at depth = 1).



LL: $\frac{1}{2}\left(1-\left(\frac{0}{1}\right)^2-\left(\frac{1}{1}\right)^2\right)=0$

RL: $\frac{1}{2}\left(1-\left(\frac{4}{9}\right)^2-\left(\frac{5}{9}\right)^2\right)=0.25$

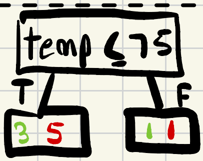
tot = $0+\frac{9}{10}(0.25)=0.225$



LL: $\frac{1}{2}\left(1-\left(\frac{2}{4}\right)^2-\left(\frac{2}{4}\right)^2\right)=0.25$

RL: $\frac{1}{2}\left(1-\left(\frac{2}{6}\right)^2-\left(\frac{4}{6}\right)^2\right)=0.22$

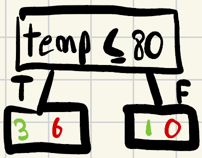
tot = $\frac{4}{10}(0.25)+\frac{6}{10}(0.22)=0.232$



LL = $\frac{1}{2}\left(1-\left(\frac{3}{8}\right)^2-\left(\frac{5}{8}\right)^2\right)=0.234$

RL = $\frac{1}{2}\left(1-\left(\frac{1}{2}\right)^2-\left(\frac{1}{2}\right)^2\right)=0.25$

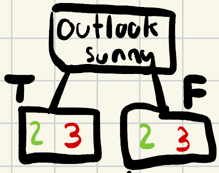
tot = $\frac{8}{10}(0.234)+\frac{2}{10}(0.25)=0.237$



LL = $\frac{1}{2}\left(1-\left(\frac{3}{9}\right)^2-\left(\frac{6}{9}\right)^2\right)=0.2$

RL = $\frac{1}{2}\left(1-\left(\frac{1}{1}\right)^2-\left(\frac{0}{1}\right)^2\right)=0$

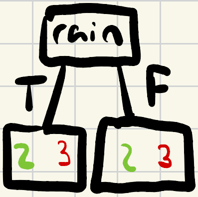
tot = $\frac{9}{10}(0.234)+0=0.21$



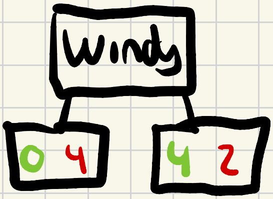
LL = $\frac{1}{2}\left(1-\left(\frac{2}{5}\right)^2-\left(\frac{3}{5}\right)^2\right)=0.24$

RL = $\frac{1}{2}\left(1-\left(\frac{2}{5}\right)^2-\left(\frac{3}{5}\right)^2\right)\ 0.24$

Tot = $\frac{5}{10}(0.24)+\frac{5}{10}(0.24)=0.24$



Same values as sunny.
its actually just one impurity

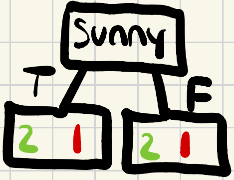


LL = $\frac{1}{2}\left(1-\left(\frac{0}{4}\right)^2-\left(\frac{4}{4}\right)^2\right)=0$

RL: $\frac{1}{2}\left(1-\left(\frac{4}{6}\right)^2-\left(\frac{2}{6}\right)^2\right)=0.22$

tot = $\frac{6}{10}(0.22)=0.132$

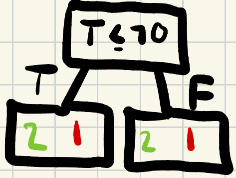
windy is the purest
Add on right side of tree



$LL = \frac{1}{2}\left(1-\left(\frac{2}{3}\right)^2-\left(\frac{1}{3}\right)^2\right) = 0.22$

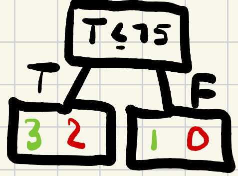
$RL = \frac{1}{2}\left(1-\left(\frac{2}{3}\right)^2-\left(\frac{1}{3}\right)^2\right) = 0.22$

$Tot = \frac{3}{6}(0.22) + \frac{3}{6}(0.22) = 0.22$



Same as Sunny

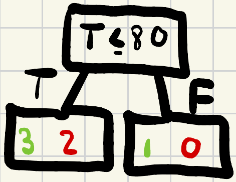
$Tot = 0.22$



$LL = \frac{1}{2}\left(1-\left(\frac{3}{5}\right)^2-\left(\frac{2}{5}\right)^2\right) = 0.24$

$RL = 0$

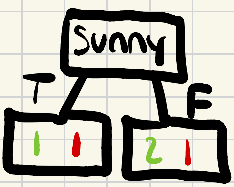
$tot = \frac{5}{6}(0.24) = 0.2$



$LL = \frac{1}{2}\left(1-\left(\frac{3}{5}\right)^2-\left(\frac{2}{5}\right)^2\right) = 0.24$

$RL = 0$

$tot = \frac{5}{6}(0.24) = 0.2$

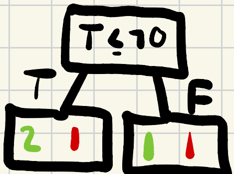


$LL = \frac{1}{2}\left(1-\left(\frac{1}{2}\right)^2-\left(\frac{1}{2}\right)^2 = 0.25\right.$

$RL = \frac{1}{2}\left(1-\left(\frac{2}{3}\right)^2-\left(\frac{1}{3}\right)^2\right) = 0.22$

$Tot = \frac{2}{5}(0.25) + \frac{3}{5}(0.22) = 0.232$

$T \leq 75$ and $T \leq 80$ are the same pick $T \leq 80$

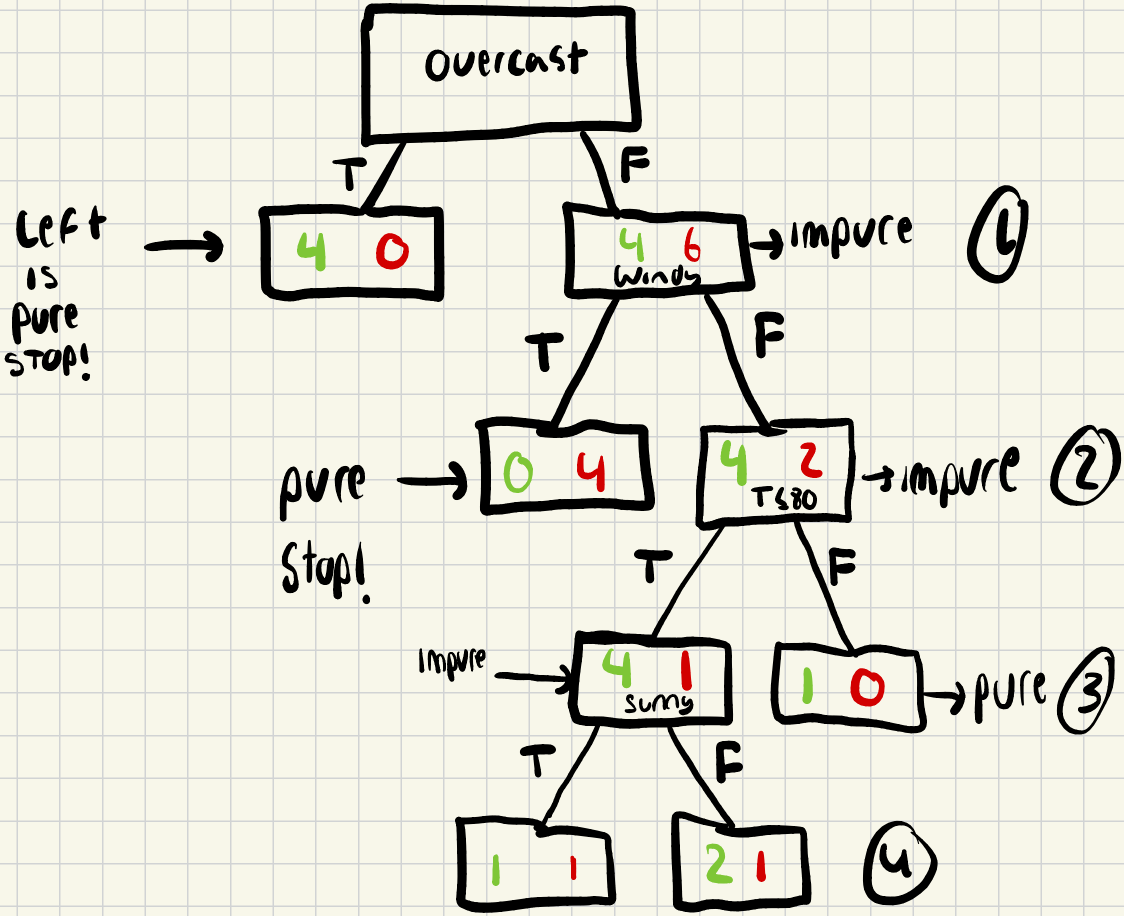


$LL = \frac{1}{2}\left(1-\left(\frac{2}{3}\right)^2-\left(\frac{1}{3}\right)^2\right) = 0.22$

$RL = \frac{1}{2}\left(1-\left(\frac{1}{2}\right)^2-\left(\frac{1}{2}\right)^2 = 0.25\right.$

$Tot = \frac{3}{5}(0.22) + \frac{2}{5}(0.25) = 0.232$

Same impurity pick eiter Sunny


Overcast
T
F
Left is pure STOP!
4 0
4 6
Windy
impure
1
T
F
pure Stop!
0 4
4 2
T≤80
impure
2
T
F
Impure
4 1
Sunny
1 0
pure
3
T
F
1 1
2 1
4

Question 2 (**20 points**) Consider the following dataset of unlabeled points:

Point	X-coordinate	Y-coordinate
P1	9	1
P2	1	1
P3	9	2
P4	8	1
P5	9	20
P6	2	2
P7	8	2
P8	1	2
P9	2	1

$$D = \sqrt{(x_2-x_1)^2 - (y_2-y_1)^2}$$

Your goal is to use K-means and K-medoids to cluster these data into two clusters. Use the Euclidean Distance as the distance measure.

(a) **(6 points)** Using P1 and P2 as the starting centroids, perform three iterations of K-means to cluster all the data points.

$k=2$ $\quad$ P1 (9,1), P2 (1,1)

	D(P1)	D(P2)
P1	$\sqrt{(9-9)^2+(1-1)^2} = 0$	$\sqrt{(1-9)^2+(1-1)^2} = 8$
P2	$\sqrt{(9-1)^2+(1-1)^2} = 8$	$\sqrt{(1-1)^2+(1-1)^2} = 0$
P3	$\sqrt{(9-9)^2+(1-2)^2} = 1$	$\sqrt{(1-9)^2+(1-2)^2} = 8.06$
P4	$\sqrt{(9-8)^2+(1-1)^2} = 1$	$\sqrt{(1-8)^2+(1-1)^2} = 7$
P5	$\sqrt{(9-9)^2+(1-20)^2} = 19$	$\sqrt{(1-9)^2+(1-20)^2} = 20.62$
P6	$\sqrt{(9-2)^2+(1-2)^2} = 7.07$	$\sqrt{(1-2)^2+(1-2)^2} = 1.41$
P7	$\sqrt{(9-8)^2+(1-2)^2} = 1.41$	$\sqrt{(1-8)^2+(1-2)^2} = 7.07$
P8	$\sqrt{(9-1)^2+(1-2)^2} = 8.06$	$\sqrt{(1-1)^2+(1-2)^2} = 1$
P9	$\sqrt{(9-2)^2+(1-1)^2} = 7$	$\sqrt{(1-2)^2+(1-1)^2} = 1$

New centroids

$$C1_x = \left(\frac{9+9+8+9+8}{5}\right) = 8.6$$

$$C1_y = \left(\frac{1+2+1+20+2}{5}\right) = 5.2$$

$C1(8.6, 5.2)$

$$C2_x = \left(\frac{1+2+1+2}{4}\right) = 1.5$$

$$C2_y = \left(\frac{1+2+2+1}{4}\right) = 1.5$$

$C2 = (1.5, 1.5)$

I2

	D(C1)	D(C2)
P1	$\sqrt{(8.6-9)^2+(5.2-1)^2} = 4.21$	$\sqrt{(1.5-9)^2+(1.5-1)^2} = 7.52$
P2	$\sqrt{(8.6-1)^2+(5.2-1)^2} = 8.68$	$\sqrt{(1.5-1)^2+(1.5-1)^2} = 0.71$
P3	$\sqrt{(8.6-9)^2+(5.2-2)^2} = 3.22$	$\sqrt{(1.5-9)^2+(1.5-2)^2} = 7.52$
P4	$\sqrt{(8.6-8)^2+(5.2-1)^2} = 4.24$	$\sqrt{(1.5-8)^2+(1.5-1)^2} = 6.52$
P5	$\sqrt{(8.6-9)^2+(5.2-20)^2} = 14.80$	$\sqrt{(1.5-9)^2+(1.5-20)^2} = 19.96$
P6	$\sqrt{(8.6-2)^2+(5.2-2)^2} = 7.33$	$\sqrt{(1.5-2)^2+(1.5-2)^2} = 0.71$
P7	$\sqrt{(8.6-8)^2+(5.2-2)^2} = 3.25$	$\sqrt{(1.5-8)^2+(1.5-2)^2} = 6.52$
P8	$\sqrt{(8.6-1)^2+(5.2-2)^2} = 8.25$	$\sqrt{(1.5-1)^2+(1.5-2)^2} = 0.71$
P9	$\sqrt{(8.6-2)^2+(5.\ -1)^2} = 7.82$	$\sqrt{(1.5-2)^2+(1.5-1)^2} = 0.71$

New centroids

$$C1_x = \left(\frac{9+9+8+9+8}{5}\right) = 8.6$$

$$C1_y = \left(\frac{1+2+1+20+2}{5}\right) = 5.2$$

C1(8.6, 5.2)

$$C2_x = \left(\frac{1+2+1+2}{4}\right) = 1.5$$

$$C2_y = \left(\frac{1+2+2+1}{4}\right) = 1.5$$

C2 = (1.5, 1.5)

C1 and C2 Stayed the Same, no need for third itteration

(b) **(10 points)** K-medoids is a variation of K-means where the cluster heads (medoids) are always chosen from the available data points. A cluster medoid, in each iteration, is given as the point with lowest sum of distances to all other points in the cluster (this is one of the several methods to compute the medoids). Using P1 and P2 as the starting medoids, perform three iterations of K-medoids to cluster all the data points.

	D(P1)	D(P2)
P1	$\sqrt{(9-9)^2+(1-1)^2} = 0$	$\sqrt{(1-9)^2+(1-1)^2} = 8$
P2	$\sqrt{(9-1)^2+(1-1)^2} = 8$	$\sqrt{(1-1)^2+(1-1)^2} = 0$
P3	$\sqrt{(9-9)^2+(1-2)^2} = 1$	$\sqrt{(1-9)^2+(1-2)^2} = 8.06$
P4	$\sqrt{(9-8)^2+(1-1)^2} = 1$	$\sqrt{(1-8)^2+(1-1)^2} = 7$
P5	$\sqrt{(9-9)^2+(1-20)^2} = 19$	$\sqrt{(1-9)^2+(1-20)^2} = 20.62$
P6	$\sqrt{(9-2)^2+(1-2)^2} = 7.07$	$\sqrt{(1-2)^2+(1-2)^2} = 1.41$
P7	$\sqrt{(9-8)^2+(1-2)^2} = 1.41$	$\sqrt{(1-8)^2+(1-2)^2} = 7.07$
P8	$\sqrt{(9-1)^2+(1-2)^2} = 8.06$	$\sqrt{(1-1)^2+(1-2)^2} = 1$
P9	$\sqrt{(9-2)^2+(1-1)^2} = 7$	$\sqrt{(1-2)^2+(1-1)^2} = 1$

C1: P1, P3, P4, P5, P7

C2: P2, P6, P8, P9

$DS(P1) = D(P1,P3) + D(P1,P4) + D(P1,P5) + D(P1,P7)$
$\quad 1 + 1 + 19 + 1.41 = 22.41$

$DS(P3) = D(P3,P1) + D(P3,P4) + D(P3,P5) + D(P3,P7)$
$\quad 1 + 1.41 + 18 + 1 = 21.41$

$DS(P4) = D(P4,P1) + D(P4,P3) + D(P4,P5) + D(P4,P7)$
$\quad 1 + 1.41 + 19.03 + 1 = 22.44$

$DS(P5) = D(P5,P1) + D(P5,P3) + D(P5,P4) + D(P5,P7)$
$\quad 19 + 18 + 19.03 + 18.03 = 74.06$

$DS(P7) = D(P7,P1) + D(P7,P3) + D(P7,P4) + D(P7,P5) = 21.44$

C1 $\downarrow$ P3 (9, 2)

$$DS(P2) = D(P2,P6) + D(P2,P8) + D(P2,P9)$$
$$1.41 + 1 + 1 = 3.41$$

$$DS(P6) = D(P6,P2) + D(P6,P8) + D(P6,P9)$$
$$1.41 + 1 + 1 = 3.41$$

$$DS(P8) = D(P8,P2) + D(P8,P6) + D(P8,P9)$$
$$1 + 1 + 1.41 = 3.41$$

$$DS(P9) = D(P9,P2) + D(P9,P6) + D(P9,P8)$$
$$1 + 1 + 1.41 = 3.41$$

C2 $\downarrow$ P6 (2,2)

	D(P3)	D(P6)
P1	$\sqrt{(9-9)^2+(2-1)^2} = 1$	$\sqrt{(2-9)^2+(2-1)^2} = 7.07$
P2	$\sqrt{(9-1)^2+(2-1)^2} = 8.06$	$\sqrt{(2-1)^2+(2-1)^2} = 1.41$
P3	$\sqrt{(9-9)^2+(2-2)^2} = 0$	$\sqrt{(2-9)^2+(2-2)^2} = 7$
P4	$\sqrt{(9-8)^2+(2-1)^2} = 1.41$	$\sqrt{(2-8)^2+(2-1)^2} = 6.08$
P5	$\sqrt{(9-9)^2+(2-20)^2} = 18$	$\sqrt{(2-9)^2+(2-20)^2} = 19.31$
P6	$\sqrt{(9-2)^2+(2-2)^2} = 7$	$\sqrt{(2-2)^2+(2-2)^2} = 0$
P7	$\sqrt{(9-8)^2+(2-2)^2} = 1$	$\sqrt{(2-8)^2+(2-2)^2} = 6$
P8	$\sqrt{(9-1)^2+(2-2)^2} = 8$	$\sqrt{(2-1)^2+(2-2)^2} = 1$
	$\sqrt{(9-2)^2+(2-1)^2} = 7.07$	$\sqrt{(2-2)^2+(2-1)^2} = 1$

C1: P1, P3, P4, P5, P7

C2: P2, P6, P8, P9

cluster didnt change.

no need for 3rd iteration

(c) **(4 points)** Compare the performance of K-means and K-medoids. Reflect on the distribution of the data points and their effect on the clustering and cluster heads in K-means and K-medoids.

As expected k-mean is heavly affected by the outliers while k-medoids is not since the medoids are taken from the dataset itself. k-mean final centroids deviates towards the outliers. we also notice that both algorithm finish in 2 itterations which mean they have similar performance.

Question 3 (15 points) Consider the following dataset for the prediction of a customer's likelihood to purchase a certain product based on historical activities.

Online Activity	Product Views	Past Purchases	Purchase Likelihood
Low	Low	None	Unlikely
Medium	Moderate	Few	Moderate
High	High	Many	Likely
Low	Low	Many	Moderate
High	Moderate	None	Likely
Medium	Low	None	Unlikely
Low	High	Few	Moderate
High	Moderate	Few	Likely
Medium	Moderate	Many	Likely
Low	Low	Few	Unlikely
High	High	None	Moderate
Medium	Low	Many	Moderate
Medium	High	Few	Moderate
High	Low	Many	Likely
Low	Moderate	None	Unlikely
Medium	High	None	?????
High	High	Few	?????
Low	Moderate	Many	?????

Your goal is to build a Naive Bayes classifier based on the labeled data, and use it to classify the new unlabeled data points.

(a) **(6 Points)** For each feature, build the frequency and likelihood tables.

Frequency Table

		Purchase		
		Unlikely	Moderate	Likely
Online Activity	Low	3	2	0
	Medium	1	3	1
	High	0	1	4

Likelihood Table

		Purchase		
		Unlikely	Moderate	Likely
Online Activity	Low	3/4	2/6	0
	Medium	1/4	3/6	1/5
	High	0	1/6	4/5

Frequency Table 2

		Purchase		
		Unlikely	Moderate	Likely
Product Views	Low	3	2	1
	~~Medium~~ moderate	1	1	3
	High	0	3	1

Likelihood Table 2

		Purchase		
		Unlikely	Moderate	Likely
Product Views	Low	3/4	2/6	1/5
	~~Medium~~ moderate	1/4	1/6	3/5
	High	0	3/6	1/5

Frequency Table 3

		Purchase		
		Unlikely	Moderate	Likely
Past Purchase	None	3	1	1
	Few	1	3	1
	Many	0	2	3

Likelihood Table 3

		Purchase		
		Unlikely	Moderate	Likely
Past Purchase	None	3/4	1/6	1/5
	Few	1/4	3/6	1/5
	Many	0	2/6	3/5

(b) **(3 Points)** Compute the prior probabilities for each class.

$P(\text{Unlikely}) = 4/15$ $P(\text{moderate}) = 6/15$ $P(\text{likely}) = 5/15$

(c) **(6 Points)** Classify each of the three new data points to the appropriate class. Show your work and the steps used to identify the suitable class.

[Medium High None ?] → moderate

P(Med | unlikely) P(high | unlikely) P(None | unlikely) P(unlikely)

(1/4) (0) (1/5) (4/15) = 0

P(Med | Moderate) P(high | Moderate) P(None | Moderate) P(moderate)

(3/6) (3/6) (1/6) (6/15) = 0.016667

P(Med | likely) P(high | likely) P(None | likely) P(likely)

(1/5) (1/5) (1/5) (5/15) = 0.002667

[High High Few ?] → Moderate

P(High | unlikely) P(high | unlikely) P(Few | unlikely) P(unlikely)

0 0 1/4 4/15 = 0

P(High | Moderate) P(high | Moderate) P(Few | Moderate) P(moderate)

1/6 3/6 3/6 6/15 = 0.016667

P(High | likely) P(high | likely) P(Few | likely) P(likely)

4/5 1/5 1/5 5/15 = 0.010667

[Low moderate many ?] $\rightarrow$ moderate

P(Low | unlikely) P(moderate | unlikely) P(many | unlikely) P(unlikely)

3/4 1/4 0 4/15 = 0

P(Low | Moderate) P(moderate | Moderate) P(many | Moderate) P(moderate)

2/6 1/6 2/6 6/15 = 0.0074

P(Low | likely) P(moderate | likely) P(many | likely) P(likely)

0 3/5 3/5 5/15 = 0

Question 4 **(15 Points)** Consider the following dataset of 5 data points. Each data point has two features (a, b) and a class label $\in \{-1, 1\}$.

Data Point	a	b	Label	$a+b$	$-a+b$
P1	1	1	-1	2	0
P2	-1	-1	-1	-2	0
P3	1	-1	1	0	-2
P4	-1	1	1	0	2
P5	0.5	-0.25	1	0.25	-0.75

w initialized to $\frac{1}{5}$

Consider the following, relatively simple, models that classify a data point to class [-1] if the condition is met, and to class [1] otherwise:

(1) $a \le 0$

(2) $b \le 0$

(3) $a + b \ge 0.5$

(4) $a + b \le -0.5$

(5) $-a + b \le 0.5$

Your goal is to use AdaBoost to find an ultimate model that combines some of these models. During the process, if two models have similar performances, then the model that comes earlier in the list above should be chosen.

(a) **(10 Points)** Perform AdaBoost for 5 steps. In each step, show all the work done (table formats are recommended).

$a \le 0$	P1: ✗ P2: ✓ P3: ✓ P4: ✗ P5: ✓	$\frac{1}{5}(1) + \frac{1}{5}(1) = 0.4$
$b \le 0$	P1: ✗ P2: ✓ P3: ✗ P4: ✓ P5: ✗	$\frac{1}{5}(1) + \frac{1}{5}(1) + \frac{1}{5}(1) = 0.6$
$a+b \ge 0.5$	P1: ✓ P2: ✗ P3: ✓ P4: ✓ P5: ✓	$\frac{1}{5}(1) = 0.2$ $\boxed{\text{Best}}$
$a+b \le -0.5$	P1: ✗ P2: ✓ P3: ✓ P4: ✓ P5: ✓	$\frac{1}{5}(1) = 0.2$
$-a+b \le 0.5$	P1: ✓ P2: ✓ P3: ✗ P4: ✓ P5: ✗	$\frac{1}{5}(1) + \frac{1}{5}(1) = 0.4$

$$a_r = \ln\left(\frac{1-\epsilon}{\epsilon}\right) = \ln\left(\frac{1-0.2}{0.2}\right) = 1.38$$

	New weights	Norm
P1	0.2	0.125
P2	$0.2 \times e^{1.38} = 0.8$	0.5
P3	0.2	0.125
P4	0.2	0.125
P5	0.2	0.125

$a \leq 0$	P1: ✗ P2: ✓ P3: ✓ P4: ✗ P5: ✓	$0.125(1)+0.125(1) = 0.25$
$b \leq 0$	P1: ✗ P2: ✓ P3: ✗ P4: ✓ P5: ✗	$0.125(1)+0.125(1)+0.125(1) = 0.375$
$a+b \geq 0.5$	P1: ✓ P2: ✗ P3: ✓ P4: ✓ P5: ✓	$0.5(1) = 0.5$
$a+b \leq -0.5$	P1: ✗ P2: ✓ P3: ✓ P4: ✓ P5: ✓	$0.125(1) = 0.125$ $\boxed{\text{Best}}$
$-a+b \leq 0.5$	P1: ✓ P2: ✓ P3: ✗ P4: ✓ P5: ✗	$0.125(1)+0.125(1) = 0.25$

$$a_1 = \ln\left(\frac{1-0.125}{0.125}\right) = 1.95$$

	New weights	Norm
P1	$0.125 \times e^{1.95} = 0.879$	0.5
P2	0.5	0.286
P3	0.125	0.071
P4	0.125	0.071
P5	0.125	0.071

$a \leq 0$	P1: ✗ P2: ✓ P3: ✓ P4: ✗ P5: ✓	$0.5(1)+0.071(1) = 0.571$
$b \leq 0$	P1: ✗ P2: ✓ P3: ✗ P4: ✓ P5: ✗	$0.5(1)+0.071(1)+0.071(1) = 0.642$
$a+b \geq 0.5$	P1: ✓ P2: ✗ P3: ✓ P4: ✓ P5: ✓	$0.286(1) = 0.286$
$a+b \leq -0.5$	P1: ✗ P2: ✓ P3: ✓ P4: ✓ P5: ✓	$0.5(1) = 0.5$
$a+b < 0.5$	P1: ✓ P2: ✓ P3: ✗ P4: ✓ P5: ✗	$0.071(1)+0.071(1) = 0.142$ best

$$a_1 = \ln\left(\frac{1-0.142}{0.142}\right) = 1.799$$

	New weights	Norm
P1	0.5	0.295
P2	0.286	0.168
P3	$0.071 \times e^{1.799} = 0.429$	0.25
P4	0.071	0.042
P5	0.429	0.25

$a \leq 0$	P1: ✗ P2: ✓ P3: ✓ P4: ✗ P5: ✓	$0.295 + 0.071 = 0.337$
$b \leq 0$	P1: ✗ P2: ✓ P3: ✗ P4: ✓ P5: ✗	$0.295 + 0.25 + 0.25 = 0.795$
$a+b \geq 0.5$	P1: ✓ P2: ✗ P3: ✓ P4: ✓ P5: ✓	0.168 **Best**
$a+b \leq -0.5$	P1: ✗ P2: ✓ P3: ✓ P4: ✓ P5: ✓	0.295
$a+b < 0.5$	P1: ✓ P2: ✓ P3: ✗ P4: ✓ P5: ✗	0.5

$$a_r = \ln\left(\frac{1-0.168}{0.168}\right) = 1.6$$

	New weights	Norm
P1	0.295	0.177
P2	$0.168 \times e^{1.6} = 0.83$	0.5
P3	0.25	0.15
P4	0.042	0.025
P5	0.25	0.15

$a \leq 0$	P1: ✗ P2: ✓ P3: ✓ P4: ✗ P5: ✓	$0.177 + 0.025 = 0.2$
$b \leq 0$	P1: ✗ P2: ✓ P3: ✗ P4: ✓ P5: ✗	0.477
$a+b \geq 0.5$	P1: ✓ P2: ✗ P3: ✓ P4: ✓ P5: ✓	0.5
$a+b \leq -0.5$	P1: ✗ P2: ✓ P3: ✓ P4: ✓ P5: ✓	0.177 **Best**
$a+b < 0.5$	P1: ✓ P2: ✓ P3: ✗ P4: ✓ P5: ✗	0.3

$$a_r = \ln\left(\frac{1-0.177}{0.177}\right) = 1.54$$

	New weights	Norm
P1	0.8256	0.5
P2	0.5	0.302
P3	0.15	0.09
P4	0.025	0.0151
P5	0.15	0.09

(b) **(2 Points)** What is the final prediction model, which combines the outcomes of the 5 steps?

$$y(x) = \text{Sign}(1.38\, y_3(x) + 1.94\, y_4(x) + 1.799\, y_5(x) + 1.6\, y_3(x) + 1.54\, y_4(x))$$

$$y(x = \text{Sign}(2.98\, y_3(x) + 3.48\, y_4(x) + 1.799\, y_5(x)$$

(c) **(3 Points)** Assess the performance of the final prediction model on the given dataset in terms of classification accuracy.

$$y(P1) = \text{Sign}(2.98(-1) + 3.48(1) + 1.799(-1)$$

$$= \text{Sign}(-1.3)$$

$$y(P1) = -1 \checkmark$$

$$y(P2) = \text{Sign}(2.98(1) + 3.48(-1) + 1.799(-1))$$

$$= \text{Sign}(-2.299)$$

$$y(P2) = -1 \checkmark$$

$$y(P3) = \text{Sign}(4.7) = 1 \checkmark$$

$$y(P4) = \text{Sign}(8.3) = 1 \checkmark$$

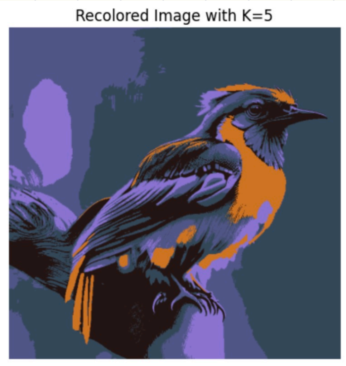
$$y(P5) = \text{Sign}(4.7) = 1 \checkmark$$

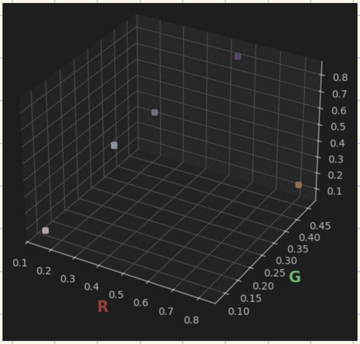
$$Acc = \frac{5}{5}$$

$= 100\%$ acc.

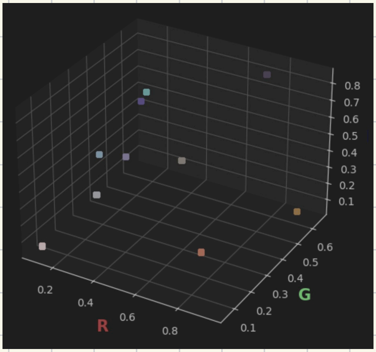
Report for Implementation Q1

K = 5: We can see that when applying K-mean algorithm with a value of 5 for K the recolored image retained the general colors but the colors we’re not as accurate as the original image. The colors were more learning towards the “blue hue”. The clustering kept 5 colors, 4 of them were some shades of blue and the last one was orange.
Giving an image that is a lot different than the original image.





K = 10: we can see that for this value of K we got a rerpresentation that is way more accurate than when it was equal to 5. IT was able to get the red colors and more variant of blue.



The error for K = 5 is bigger than the one for K = 10 (0.011 vs 0.006116) just as expected. This means that when K = 10 the recolored image was closer to the original than when K = 5.