**Dataset 1: Chocolates** 

ID	Cocoa	Shape	Tree	Price	Review	Seller	Found	Found	# in	Color
	(%)		nuts?	(\$/oz)	(stars)		online	in-store	bag	
1	30	Round	0 (No)	0.25	3	Hershey	1	0	40	Milk
2	15	Square	0	0.5	3	H & D	0	1	15	White
3	20	Square	0	0.44	2	Neuhaus	1	0	20	White
4	55.5	Square	0	0.8	4	Mars	1	0	12	Milk
5	60	Round	1 (Yes)	1.3	4	Ferrero	0	1	6	Dark
6	5.5	Round	0	1.1	3	Lindt	0	1	2	White
7	10	Square	1	0.75	5	H & D	0	1	8	White
8	70	Square	0	0.9	1	Hershey	0	1	32	Milk
9	85	Square	0	1.8	2	Mars	1	0	14	Dark
10	60	Round	1	2.4	5	Lindt	0	1	1	Dark
11	14.5	Square	0	1.5	4	Lindt	1	0	6	Dark
12	77	Round	0	1.4	3	H & D	0	1	12	Dark

Q: Given Dataset 1, plot the samples using

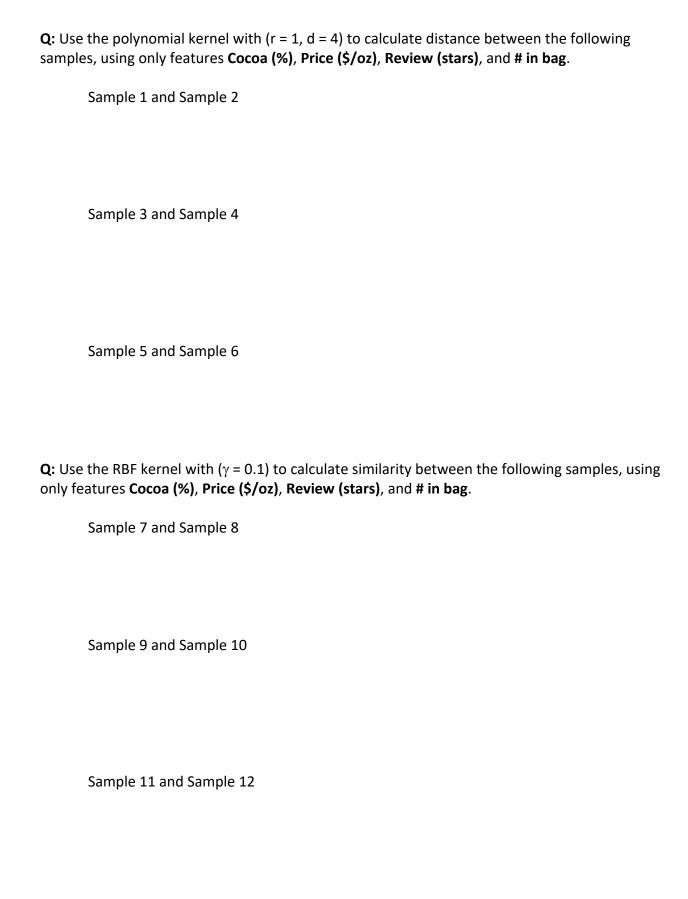
Price along the x-axis and

Review (stars) along the y-axis.

Draw a hard margin boundary to separate the positive class (Color = Dark) from the negative class (Color != Dark).

Draw the boundary and the margins.

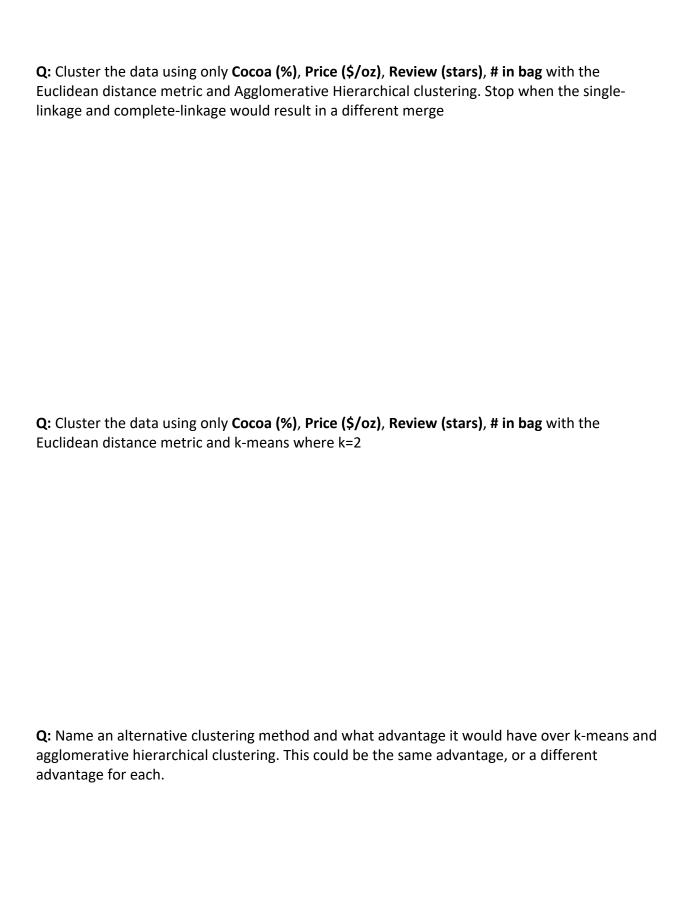
If a hard margin boundary does not exist, draw a soft margin boundary that prioritizes accuracy over margin width.



<b>Q:</b> Use Manhattan distance as your distance metric, k-nearest neighbors with k=5, and the features <b>Cocoa (%)</b> and <b># in bag</b> , for classification.  Predict <b>Color</b> of a sample with 45% Cocoa and 10 pieces in the bag
Q: Use Manhattan distance as your distance metric, k-nearest neighbors with k=3, and the features Cocoa (%) and # in bag, for regression.  Predict Price (\$/oz) of a sample with 45% Cocoa and 10 pieces in the bag

Q: Use the feature Review (stars) to build the best Decision tree with a single split. Predict Color of a sample with a 2-star review
Q: Use the feature Review (stars) to build the best Decision tree with a single split. Predict Price (\$/oz) of a sample with a 4-star review
Q: What would be the next step if I plan on making a Random Forest model? What about an Adaboost model?

Q: Train a Naïve Bayes Classifier to predict <b>Shape</b> from <b>Seller</b> and <b>Color</b>
Predict the <b>Shape</b> of a sample with Color = White and Seller = H & D
Q: What differentiates PCA from LDA? In what situation can you use PCA, but cannot use LDA?



**Dataset 2: Grocery store transactions** 

ID	Itemset					Support
1	Bread					80%
2	Milk					35%
3	Leafy vegetables		25%			
4	Root vegetables					25%
5	Bread	Milk			30%	
6	Bread	Root vegetables			20%	
7	Root vegetables	Milk			15%	
8	Bread	Leafy vegetables			5%	
9	Leafy vegetables	Milk			5%	
10	Root vegetables	Leafy vegetables			5%	
11	Root vegetables			Milk	10%	
12	Leafy vegetables			Milk	3%	
13	Root vegetables	etables		Bread	2%	
14	Root vegetables	etables		Milk	1%	
15	Leafy vegetables	vegetables Root vege		Bread	Milk	1%

**Q:** Find all informative rules from this dataset where itemsets are frequent if Support >= 10%, and rules are strong if Confidence >= 55%