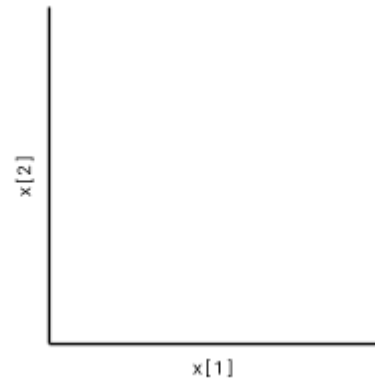
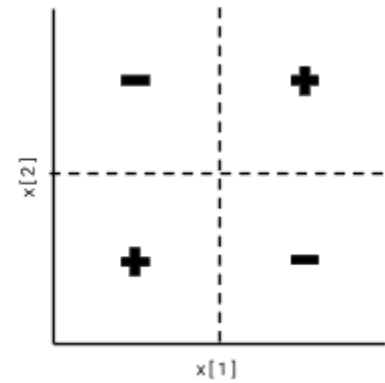
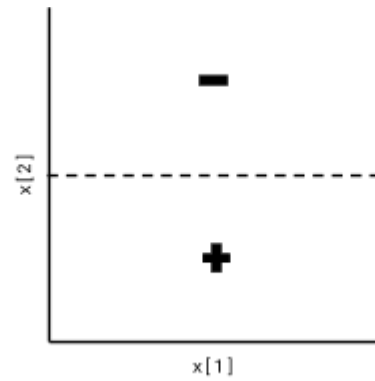
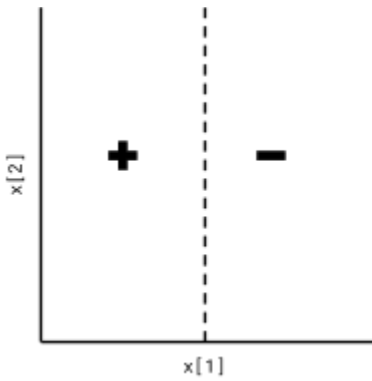


Question 1 (Random Forests)

Given the decision boundaries for three decision trees, write the decision boundary that would be created by using a random forest.



Question 2 (Boosting)

Find next iteration in Adaboost, in predicting whether it is safe or risky to provide credit given features Income, Credit history, IsSavings > \$100K, Market conditions.

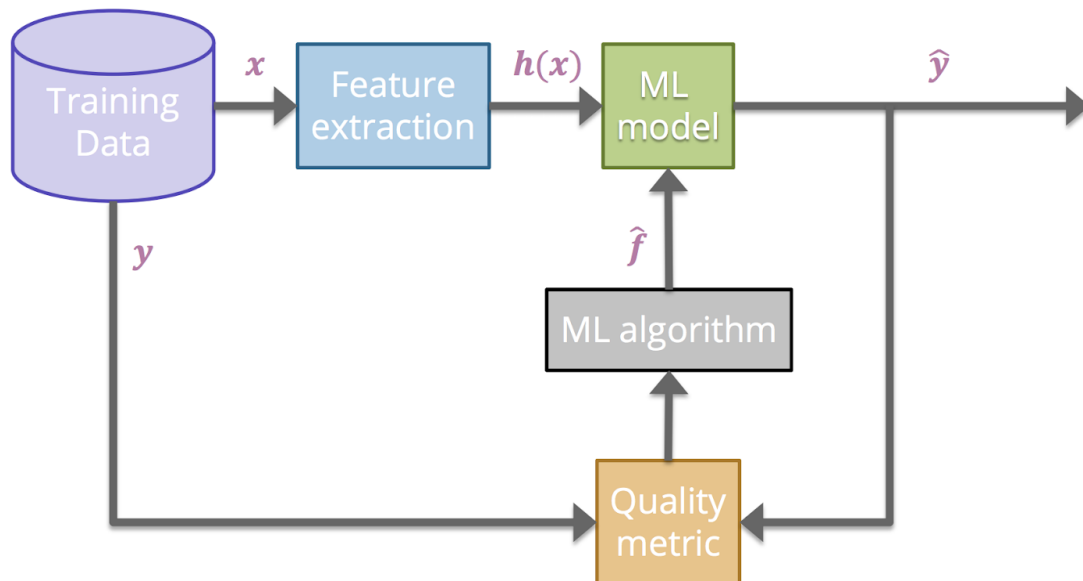
Weighted error for features : Income is 0.2, Credit history is 0.35, IsSavings > 100K is 0.35, Market conditions 0.4. Please fill in the following table the new weights by Adaboost.

Income (thousand dollars)	y	\hat{y}	Previous weight	New weight
130	Safe	Safe	0.5	
80	Risky	Risky	1.5	
110	Risky	Safe	1.5	
110	Safe	Safe	2	
90	Safe	Risky	1	
120	Safe	Safe	2.5	
30	Risky	Risky	3	
60	Risky	Risky	2	
95	Safe	Risky	0.5	
60	Safe	Risky	1	
98	Safe	Risky	0.5	

Question 3 (ML Pipeline Review)

For all the models we've looked at so far (Linear Regression, Ridge/LASSO Regression, Logistic Regression, and Decision Trees), write down the following:

1. Training Data: What does the training data consist of? What types of data can each model use?
2. Feature Extraction: What types of features can we extract for the different models, what methods have we talked about for feature extraction.
3. ML Model: What are the parameters/weights we learn?
4. ML Algorithm: What methods are used to compute the model?
5. Quality Metric: What metrics do we use to measure our models?



Compare and contrast the two classification methods we've discussed so far: Decision Trees and Logistic Regression. Some questions to get you started:

1. What features does each method use?
2. How do the quality metrics compare?
3. How do each handle continuous and categorical variables?
4. How do each prevent overfitting?