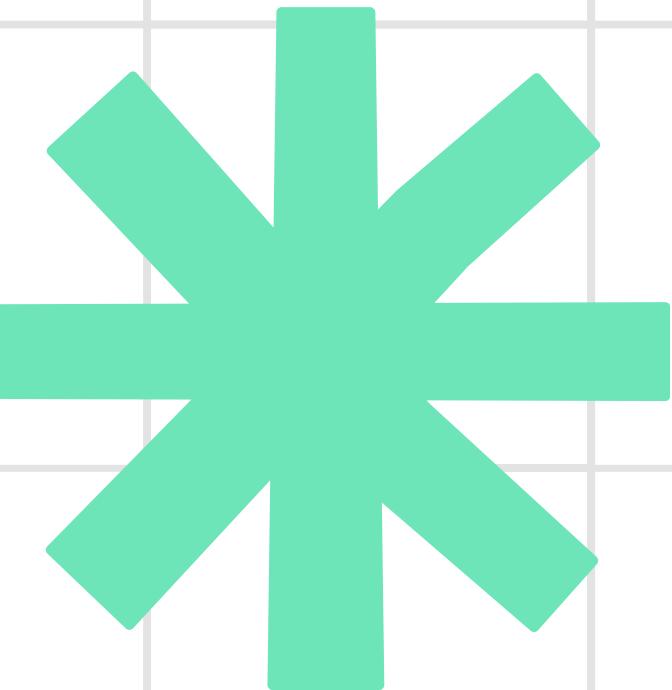


Session 2

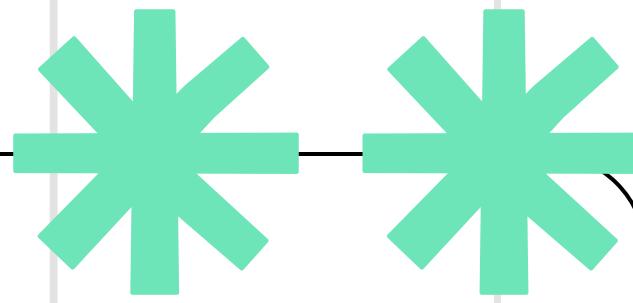
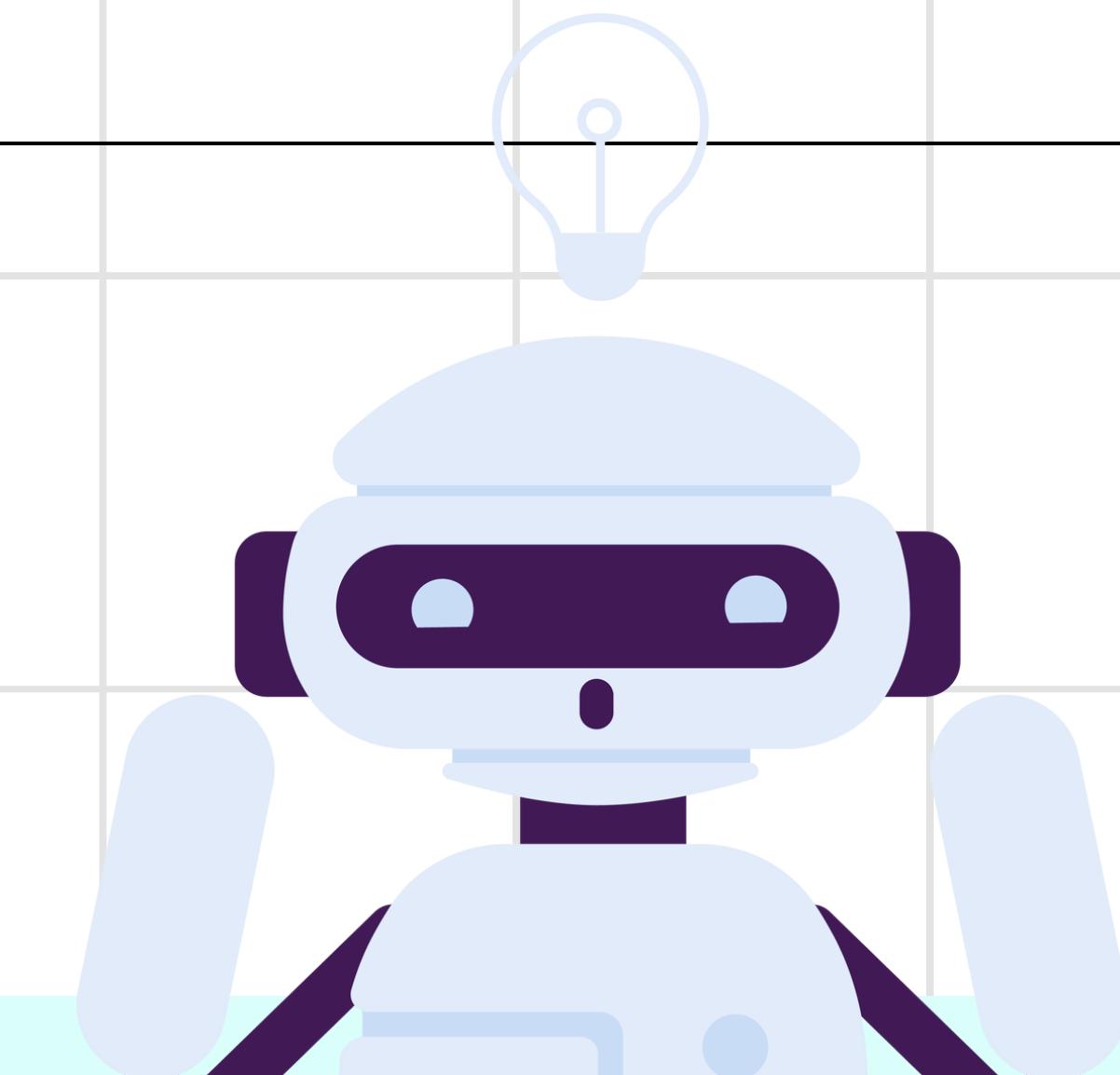
Supervised Learning

Selma MANI

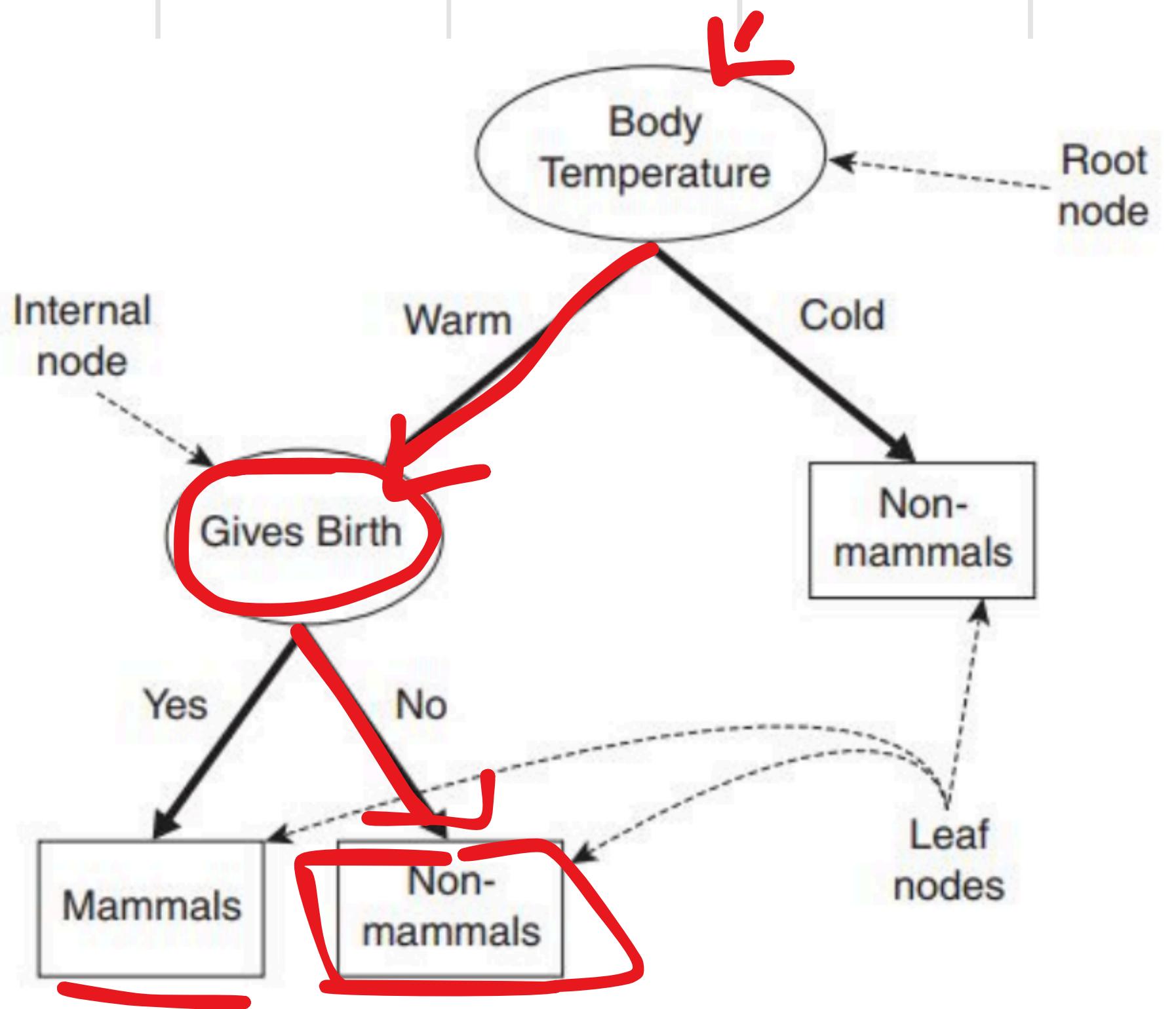


Link to the notebook we will
be using for this session ^.^

Supervised Learning Notebook



Decision Trees

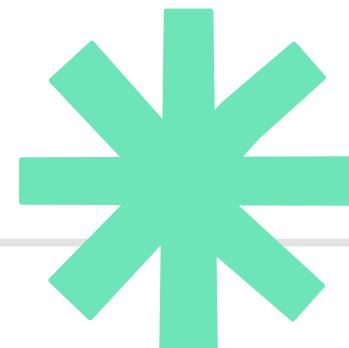


Mammals Classification Tree

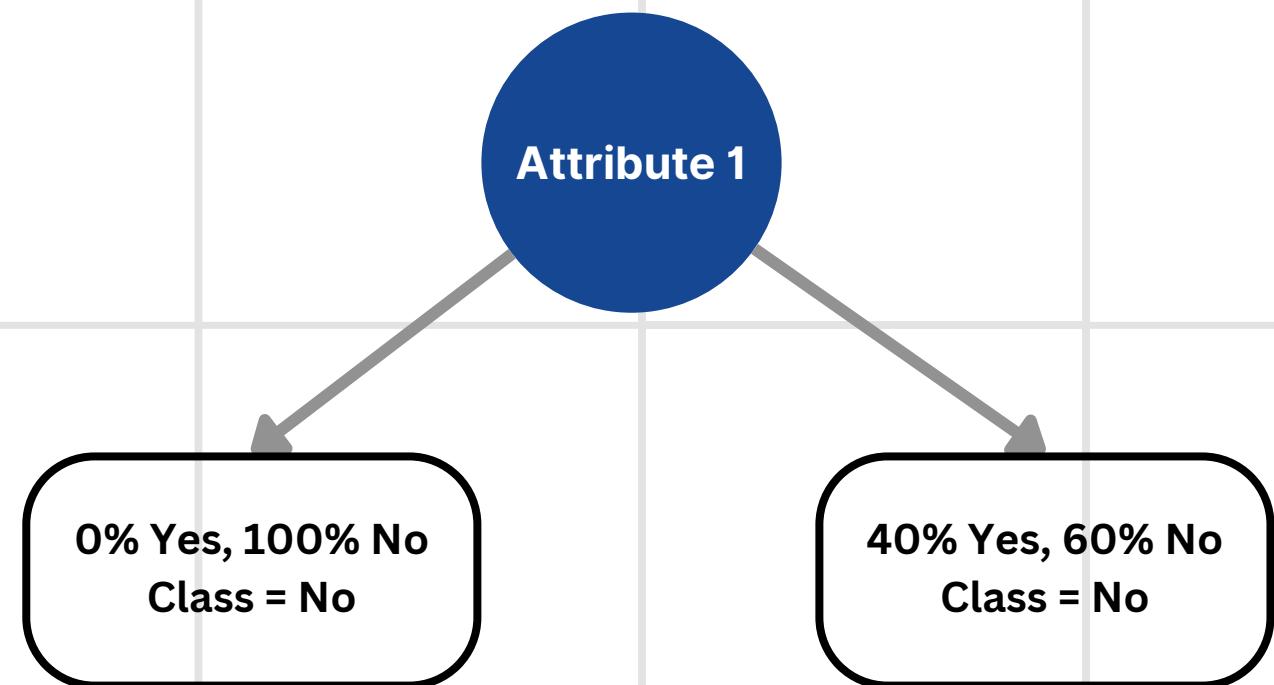
ANIMAL	BODY TEMPERATURE	GIVES BIRTH	...	CLASS
Chicken	Warm	No	?



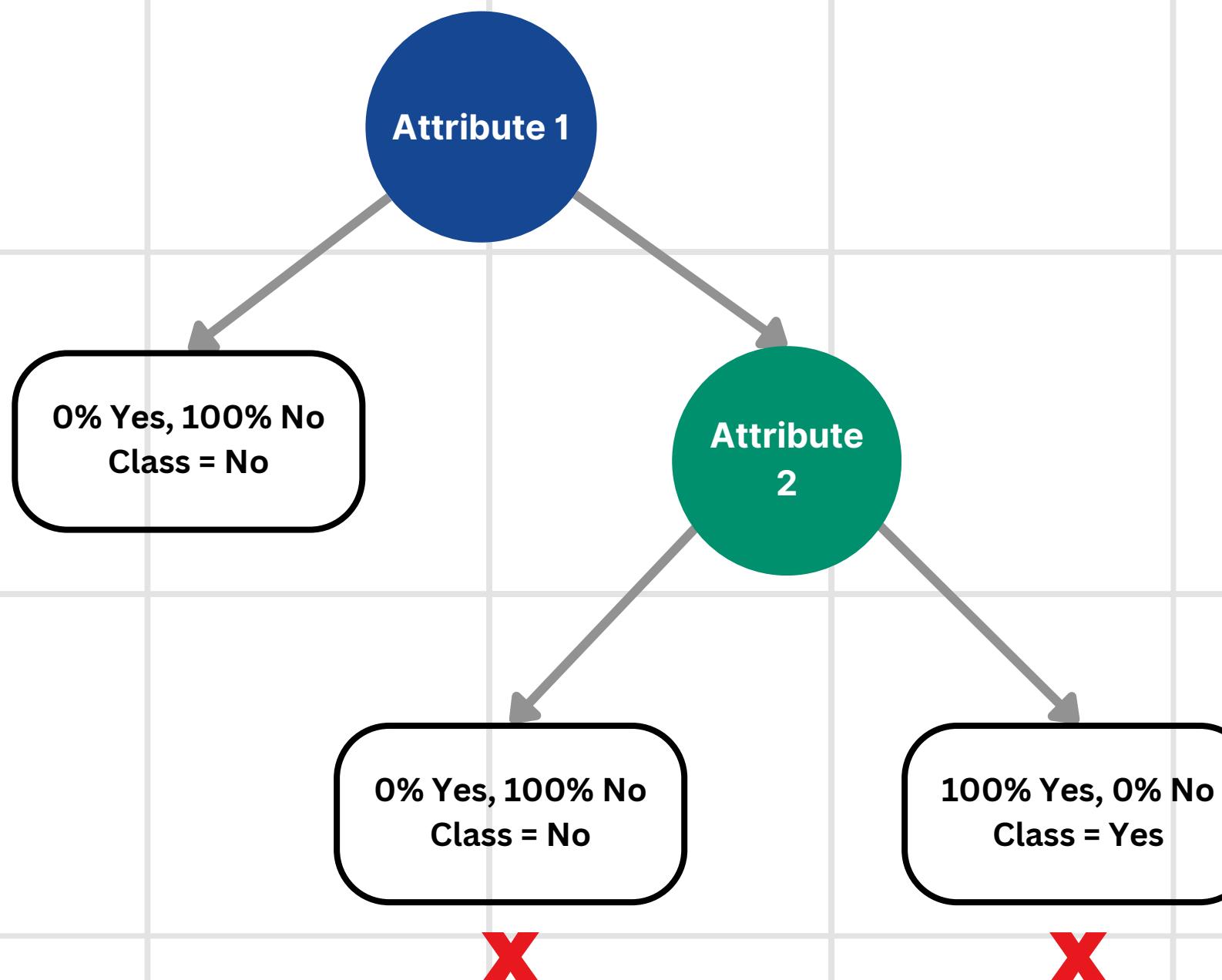
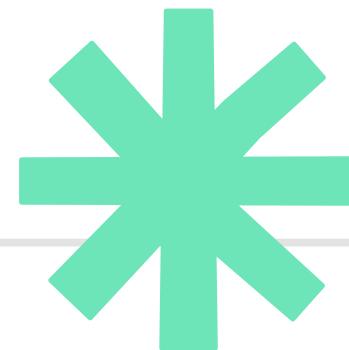
Hunt's Algorithm for decision tree building



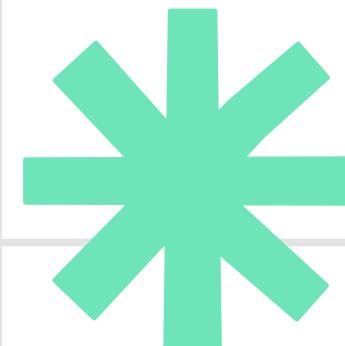
Pure



Hunt's Algorithm for decision tree building

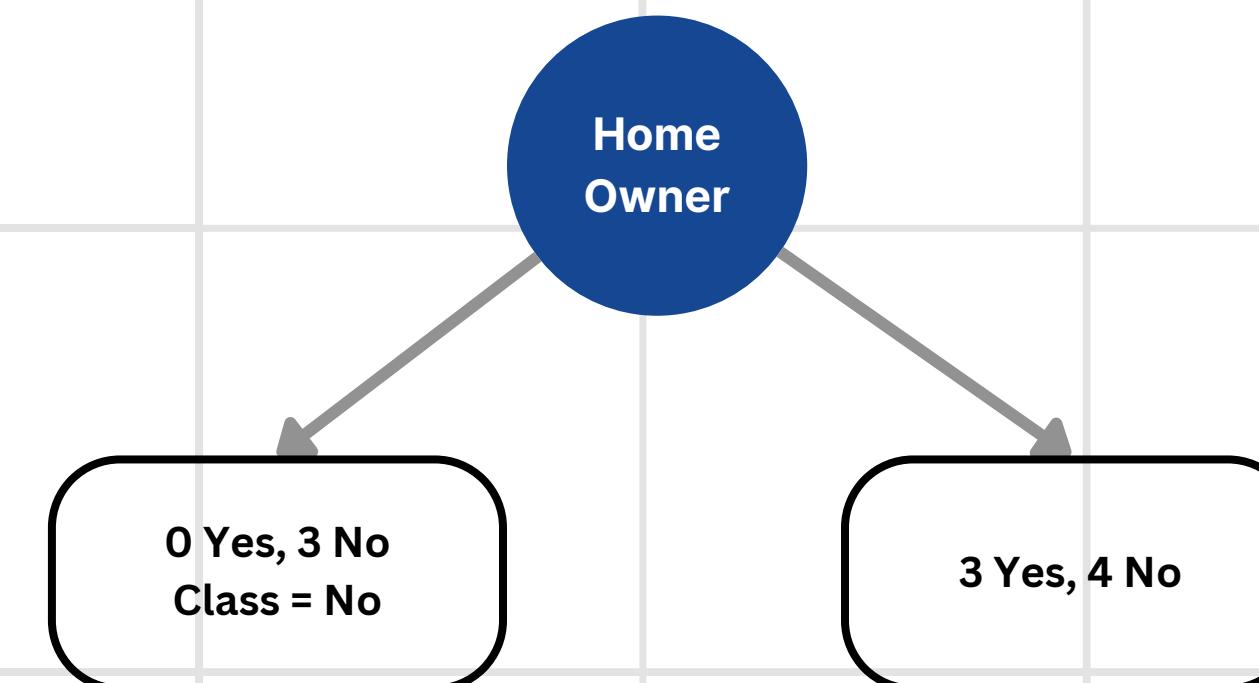


Hunt's Algorithm for decision tree building

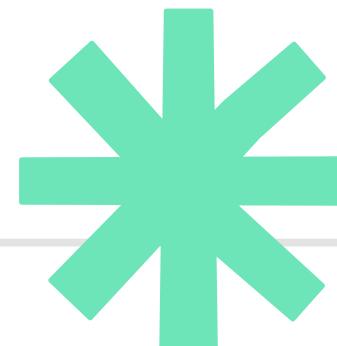


ID	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

10 rows (3 Yes, 7 No)

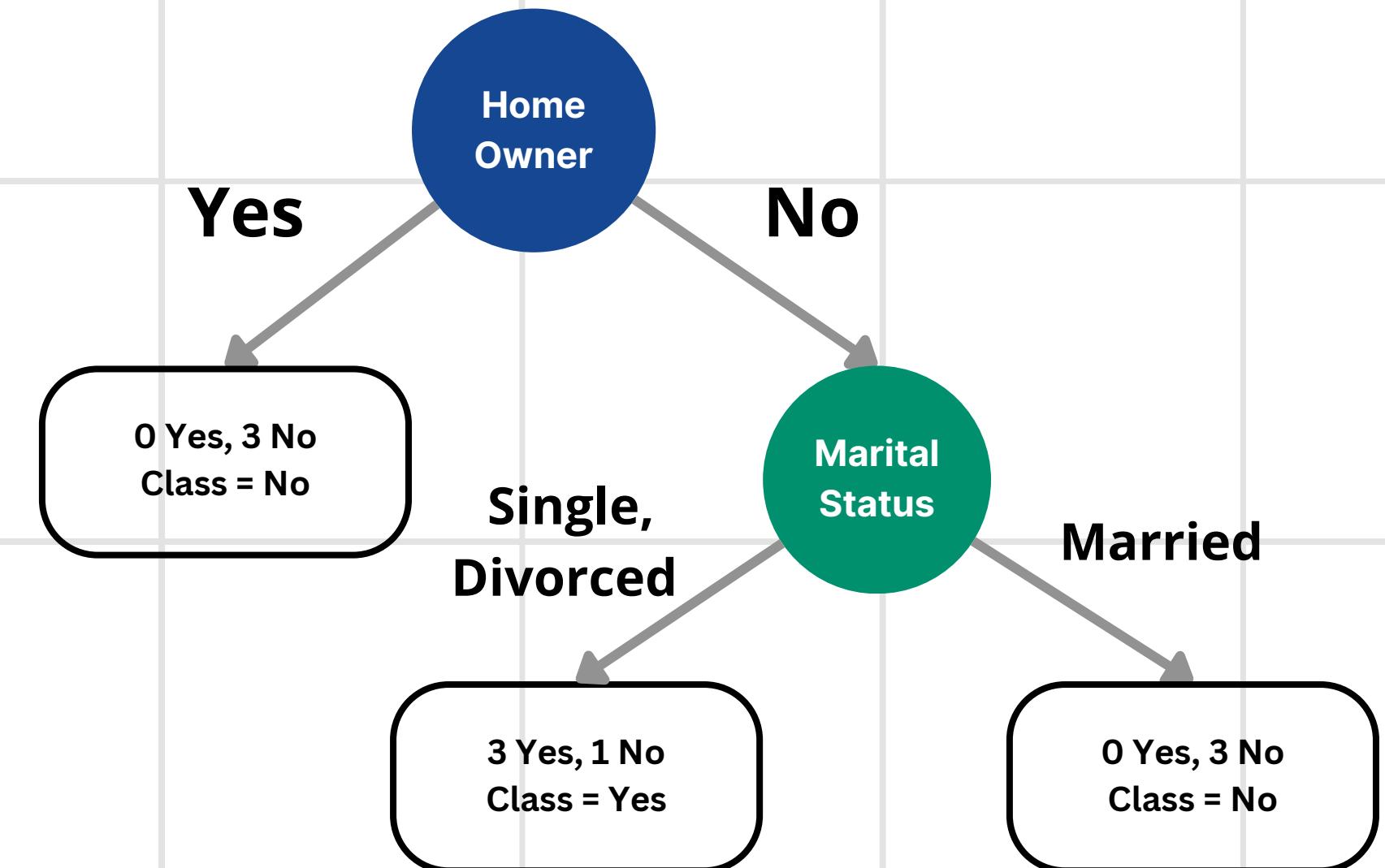


Hunt's Algorithm for decision tree building

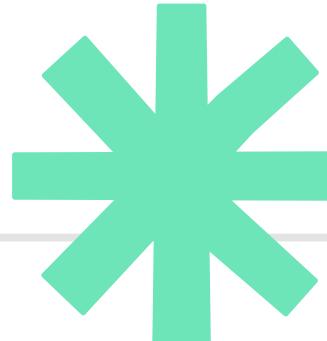


ID	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

10 rows (3 Yes, 7 No)

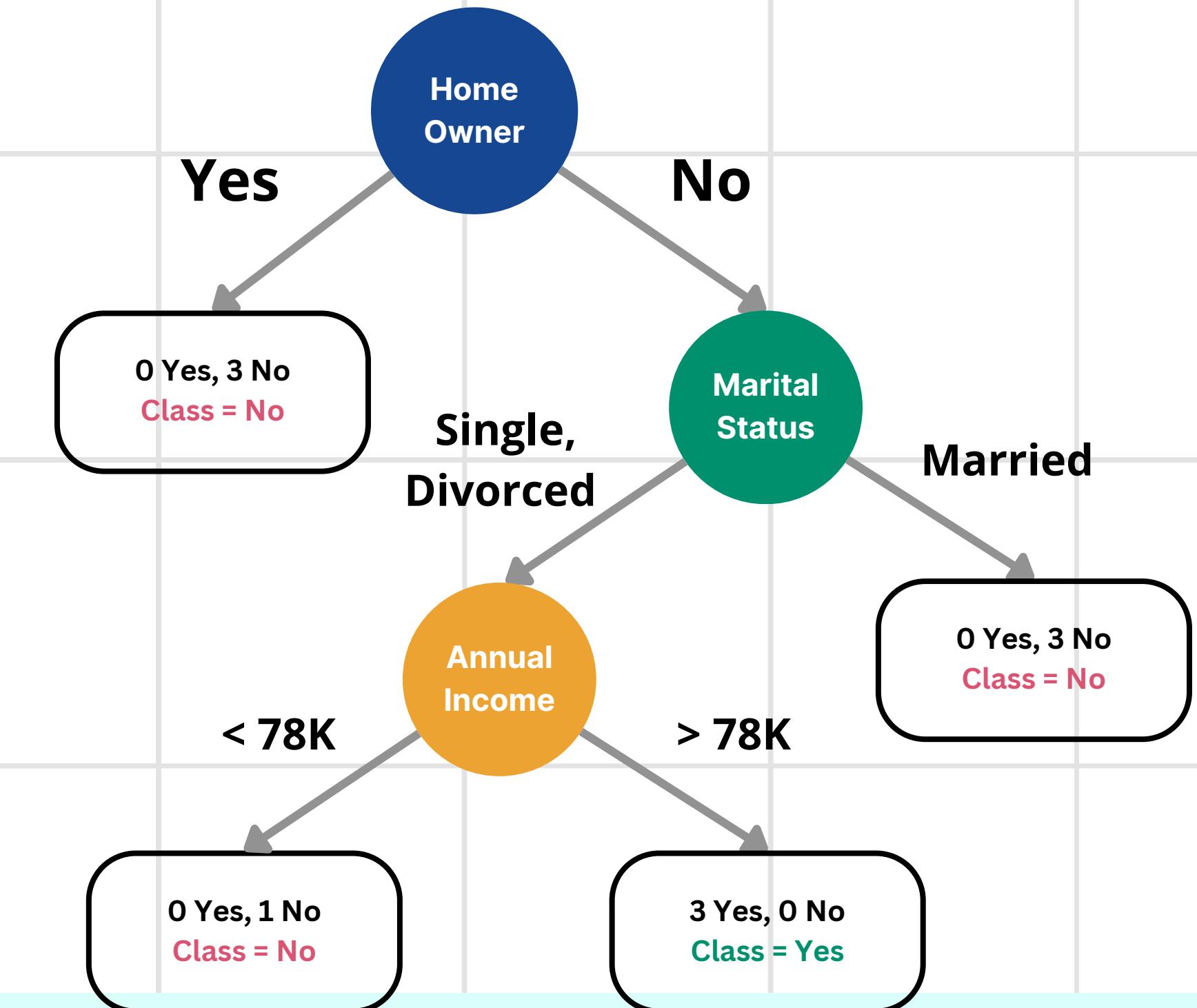


Hunt's Algorithm for decision tree building

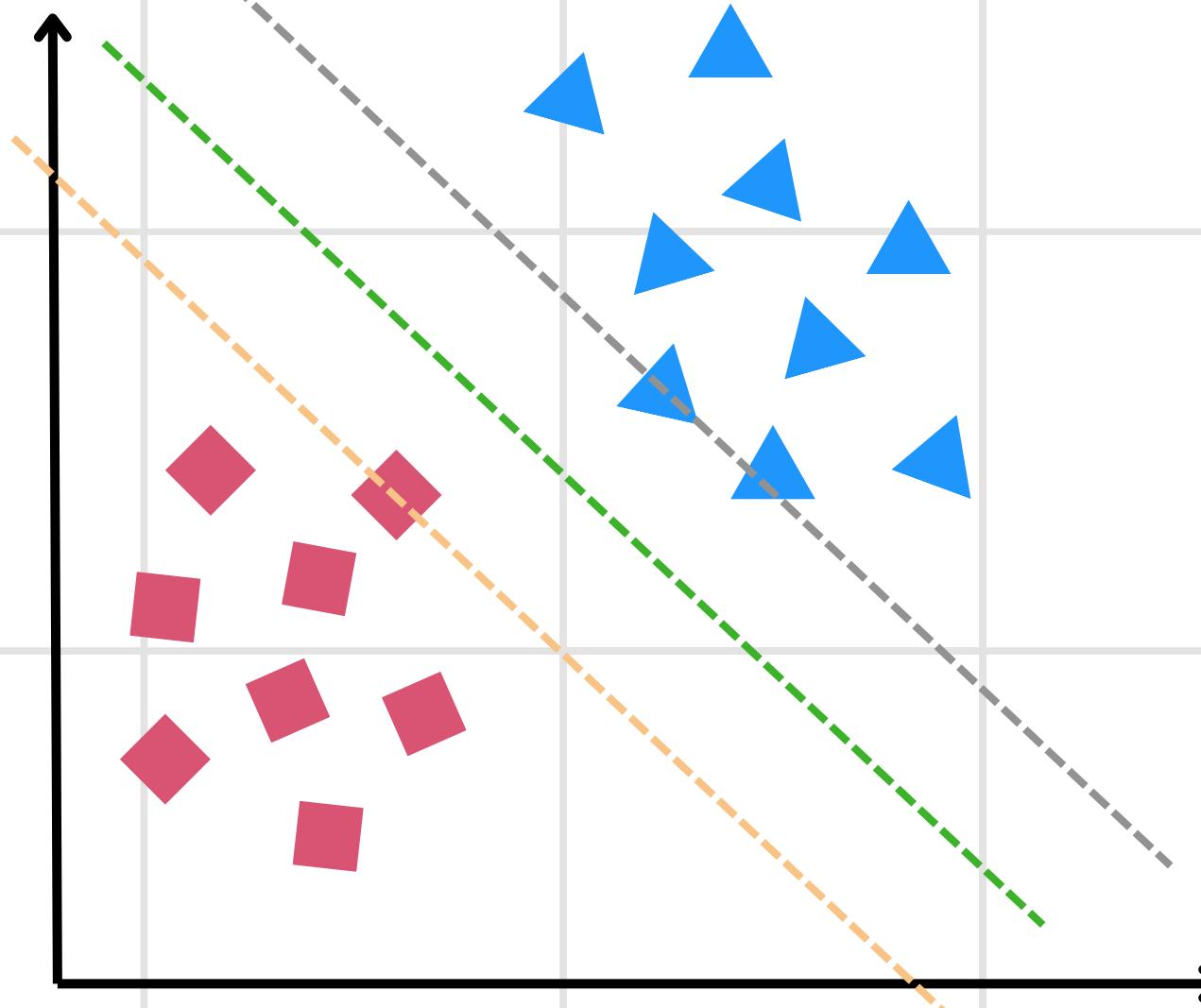
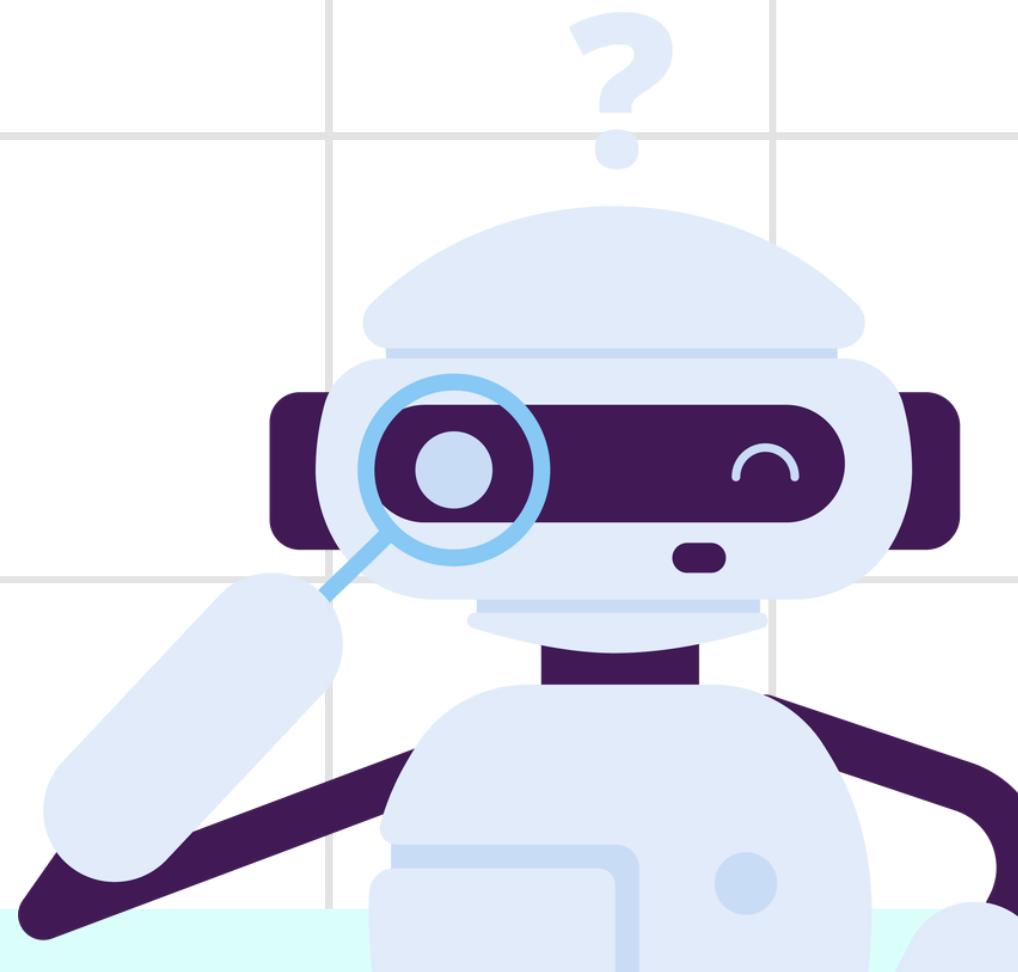
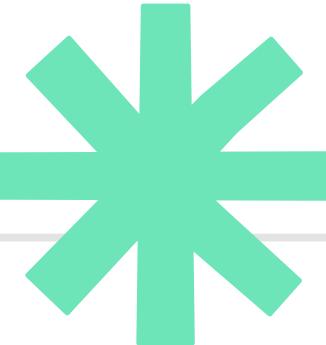


ID	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

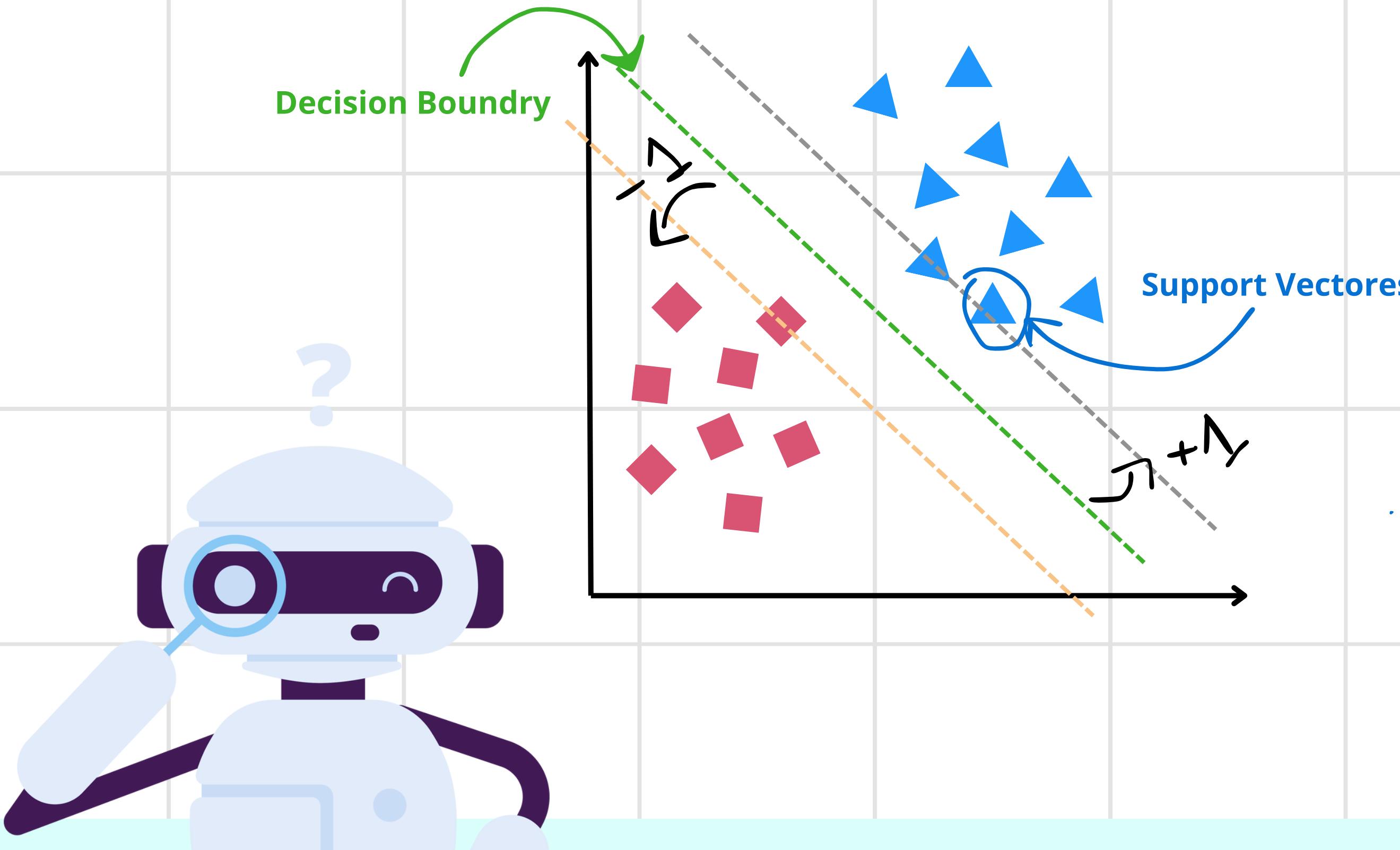
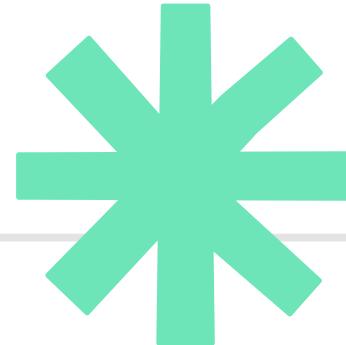
10 rows (3 Yes, 7 No)



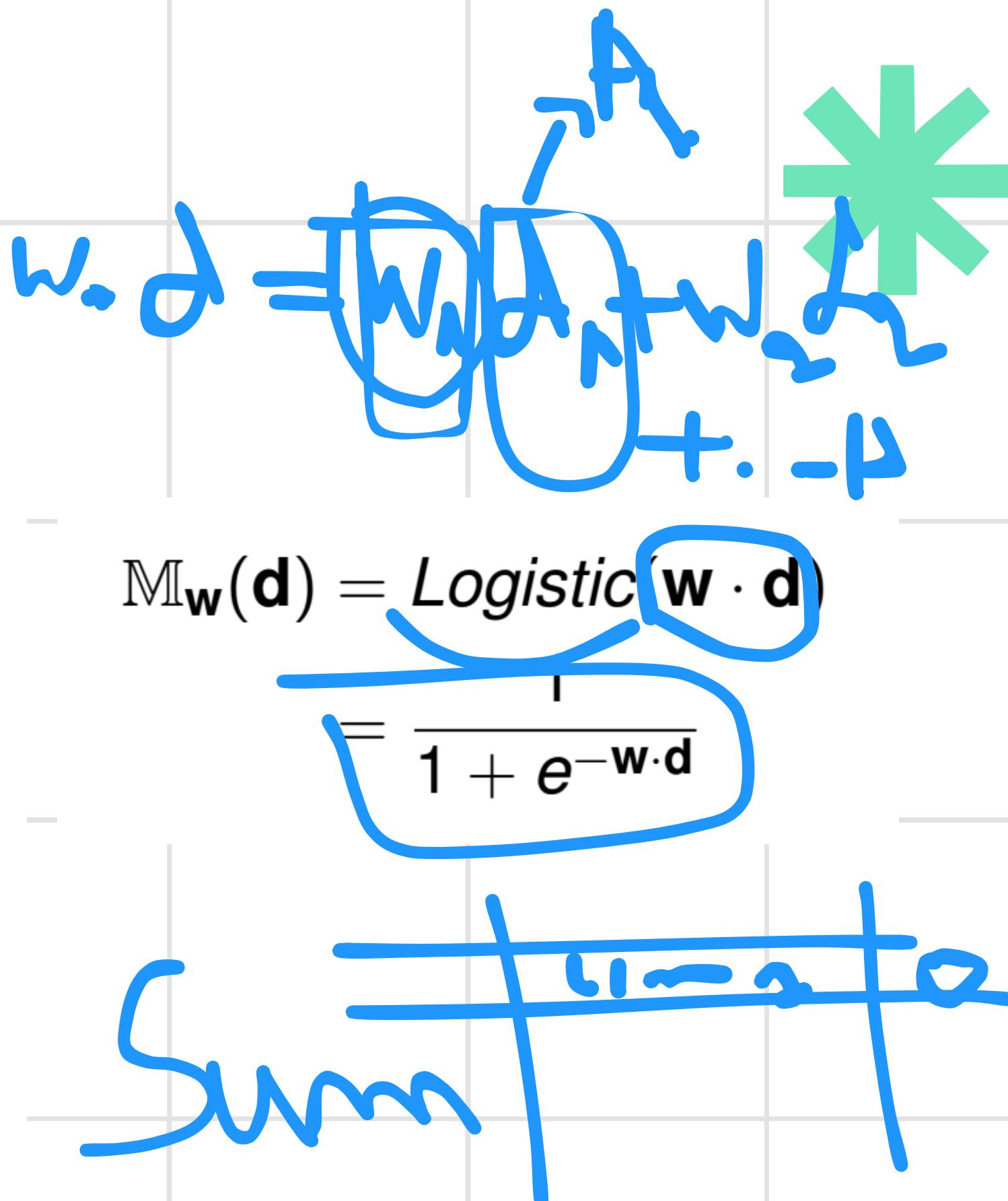
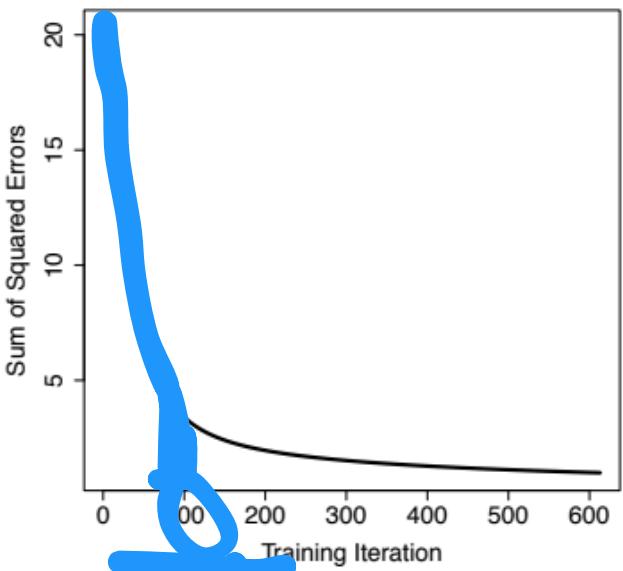
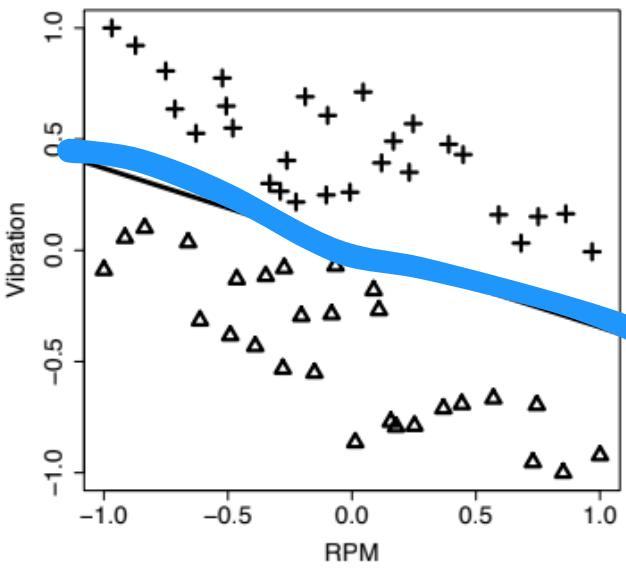
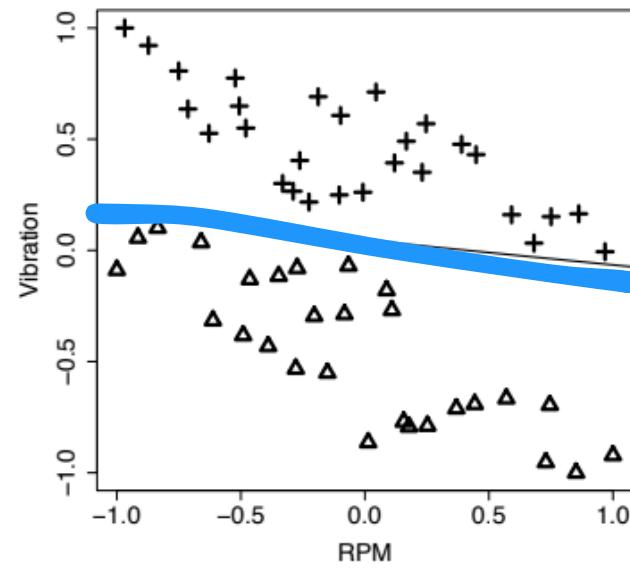
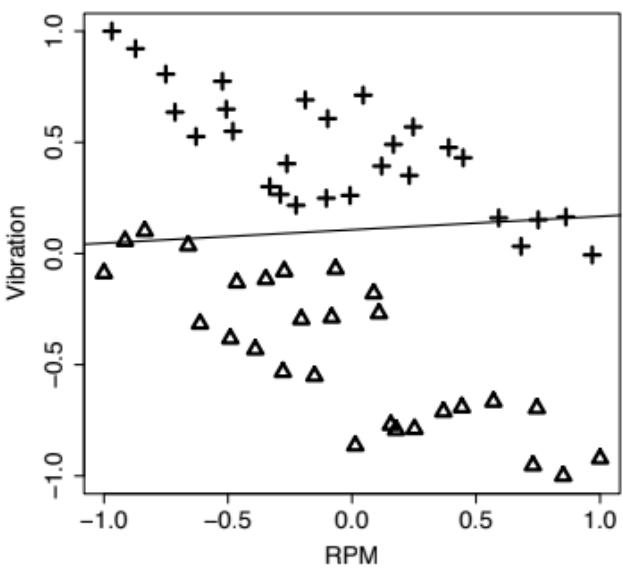
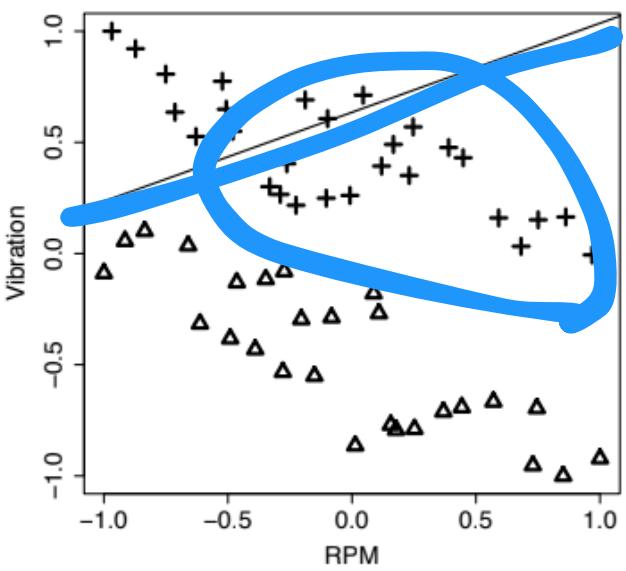
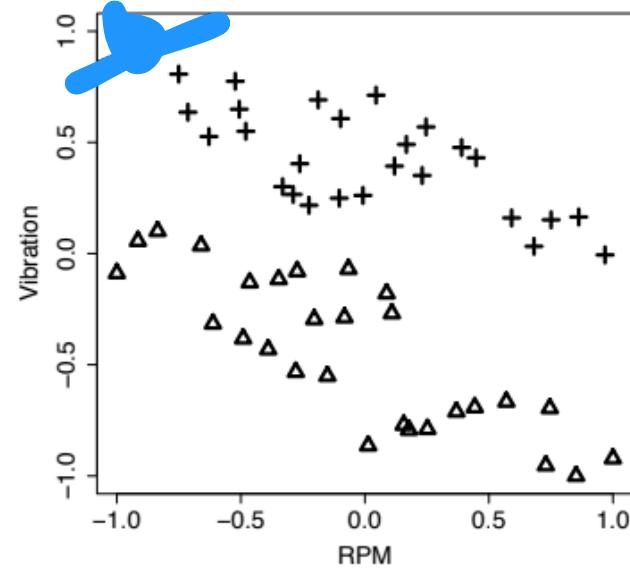
Support Vector Machines



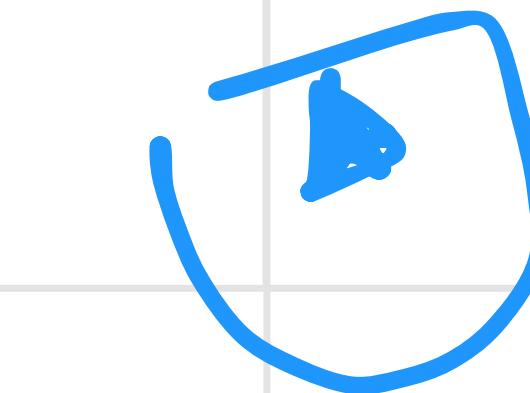
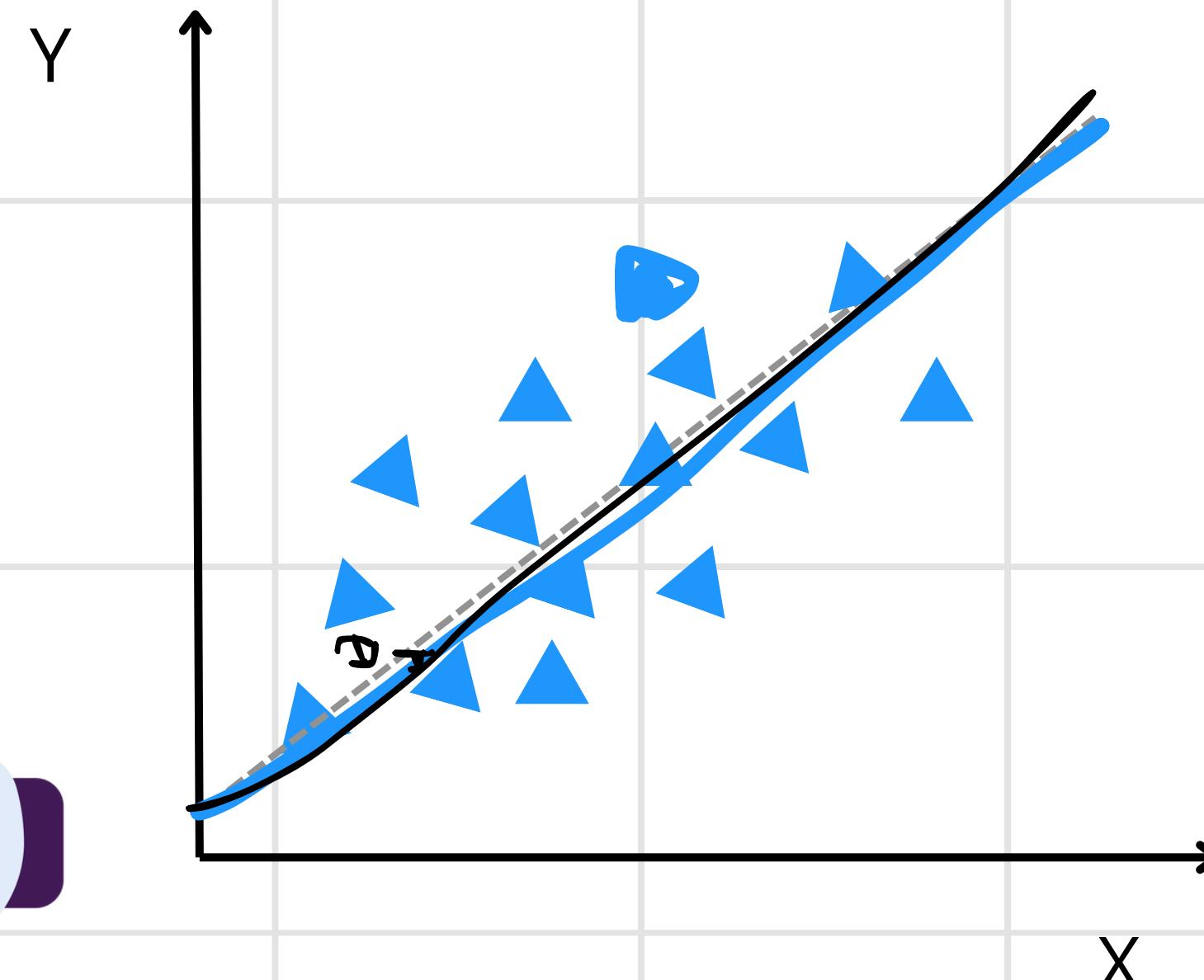
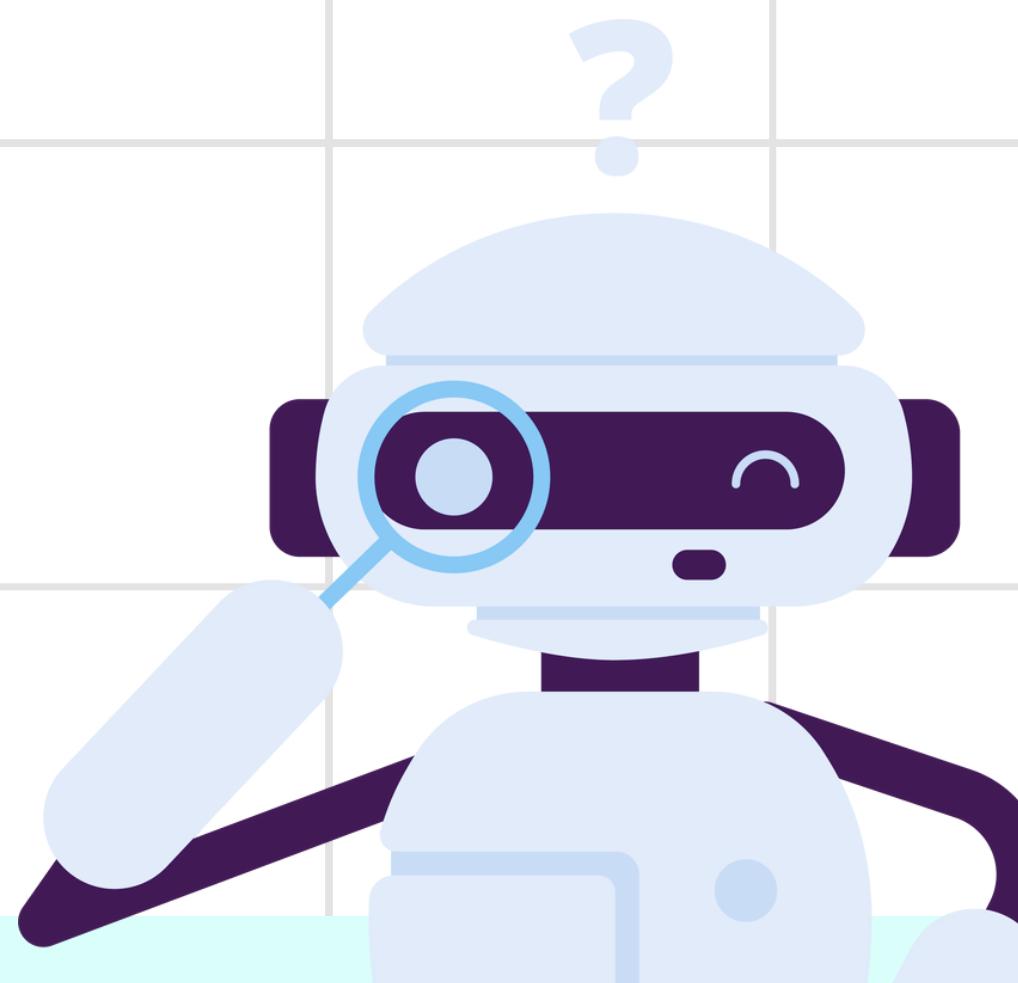
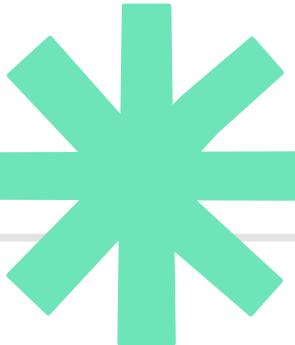
Support Vector Machines



Logistic Regression



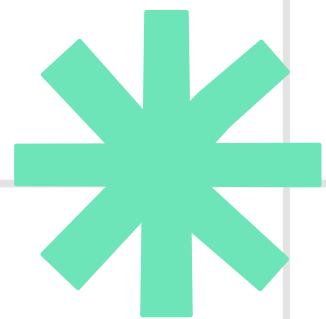
Linear Regression



Simply it is Finding a
linear function for prediction



Evaluation Metrics



Accuracy

is a metric used to evaluate the performance of a classification model.

Accuracy = Number of Correct Predictions / Total Number of Predictions

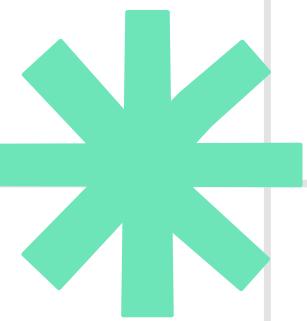
$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Confusion Matrix

is a table used to evaluate the performance of a classification model. It provides a detailed breakdown of the model's predictions compared to the actual outcomes.

		Positive	Negative
Target	Positive	TP (True Positive)	FN (False Negative)
	Negative	FP (False Positive)	TN (True Negative)
			Prediction

Performance Measures



$$\text{TPR} = \frac{TP}{(TP + FN)}$$

$$\text{TNR} = \frac{TN}{(TN + FP)}$$

$$\text{FPR} = \frac{FP}{(TN + FP)}$$

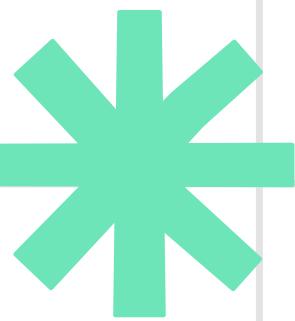
$$\text{FNR} = \frac{FN}{(TP + FN)}$$

$$\text{precision} = \frac{TP}{(TP + FP)}$$

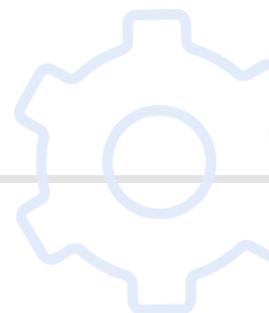
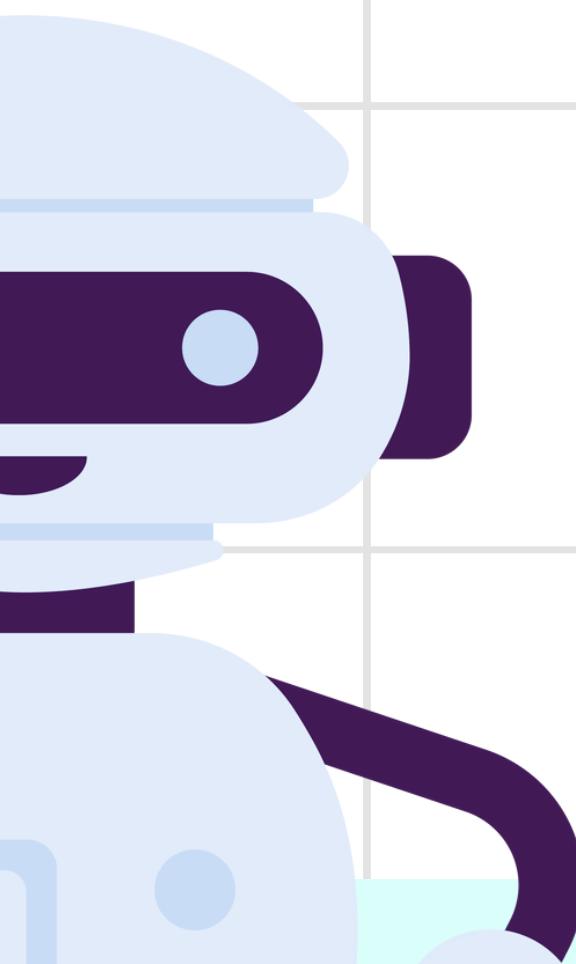
$$\text{recall} = \frac{TP}{(TP + FN)}$$

$$F_1\text{-measure} = 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})}$$

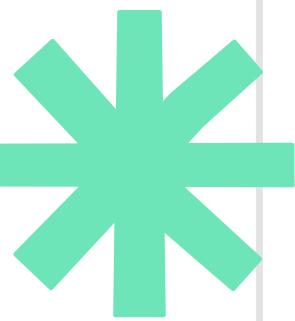
Imbalanced Dataset



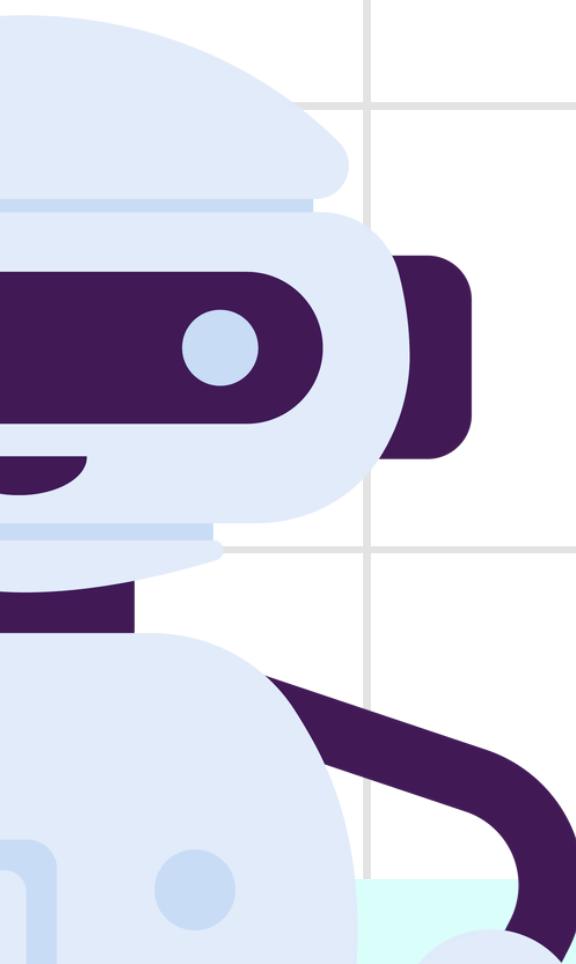
Oversampling ? Undersampling ??

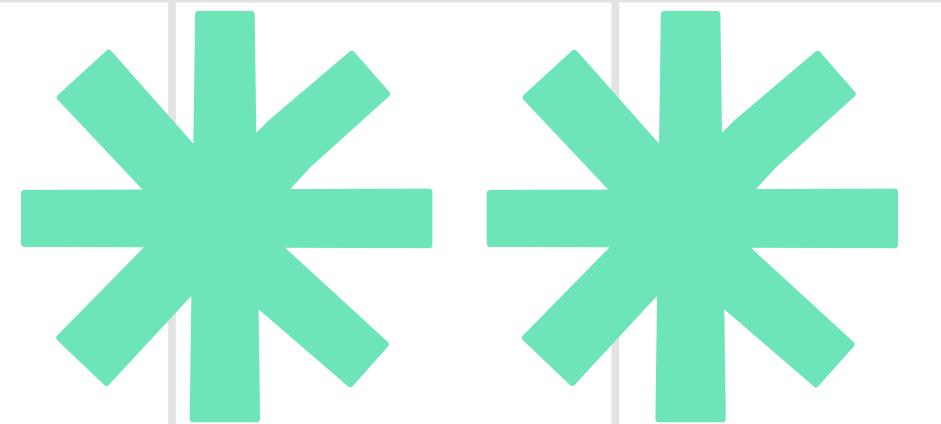


What to do Next?



READ AND UNDERSTAND THE CONCEPTS





Thank you!

I hope you
enjoyed 

