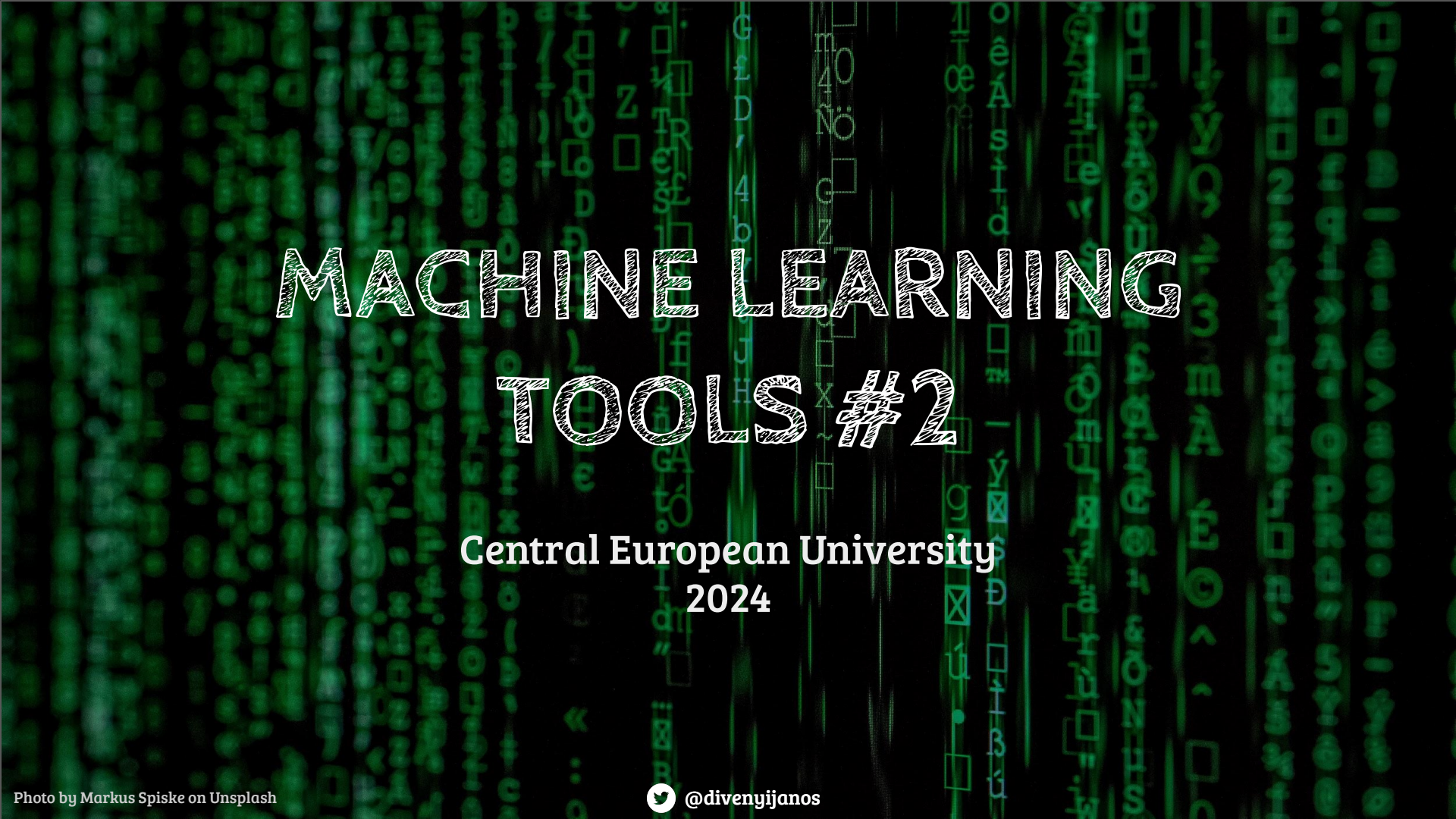


MACHINE LEARNING TOOLS #2

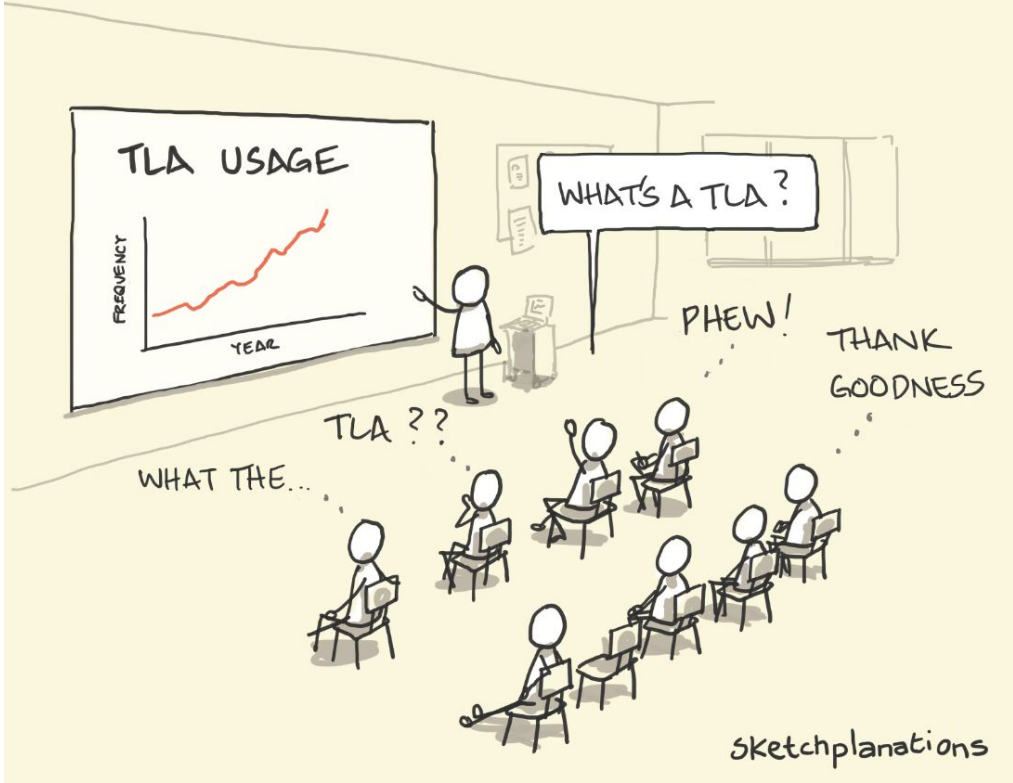
Central European University
2024

Photo by Markus Spiske on Unsplash

@divenyijanos

ASK THE QUESTION AT TALKS

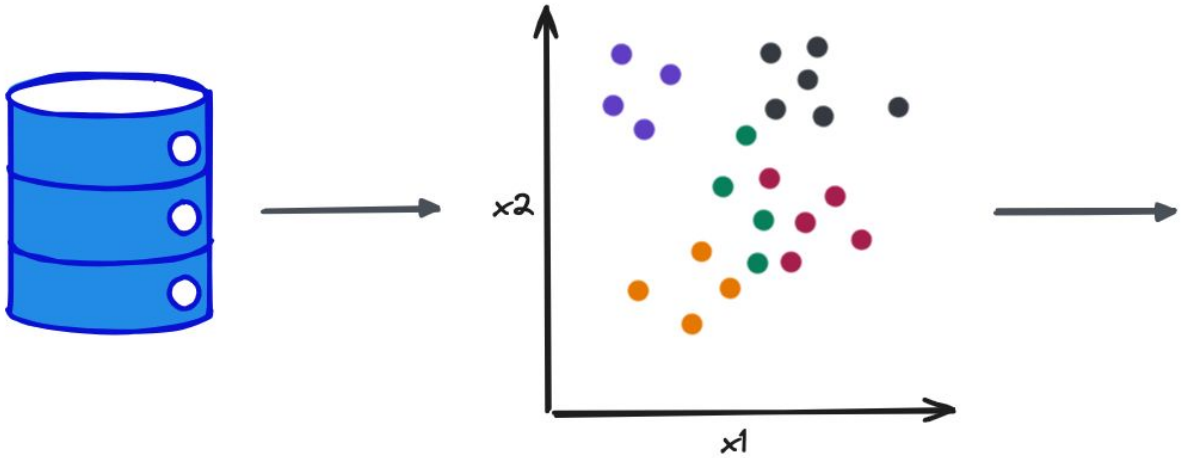
IF YOU'RE WONDERING, LIKELY OTHERS ARE TOO

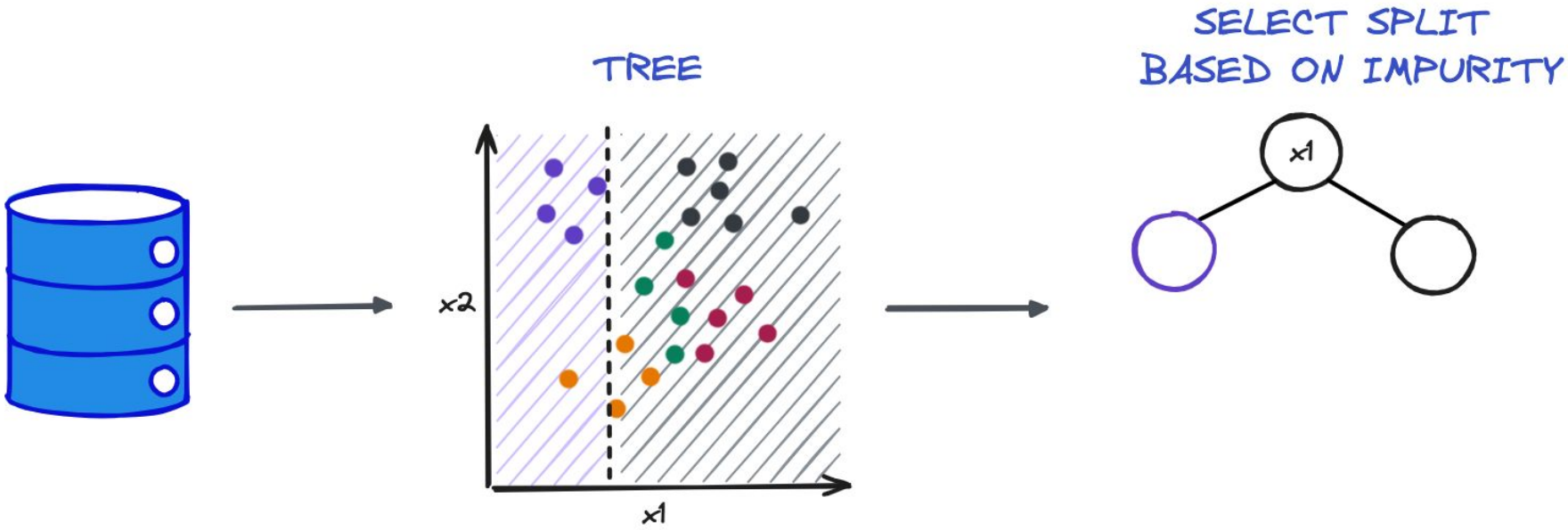


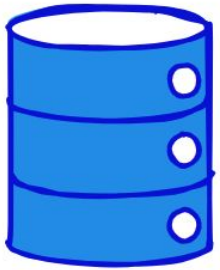
@divenyijanos

TREE

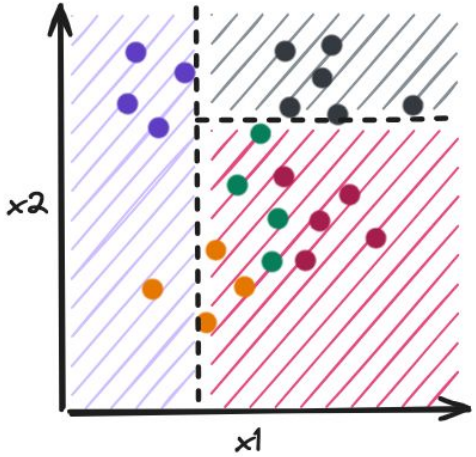
SELECT SPLIT
BASED ON IMPURITY



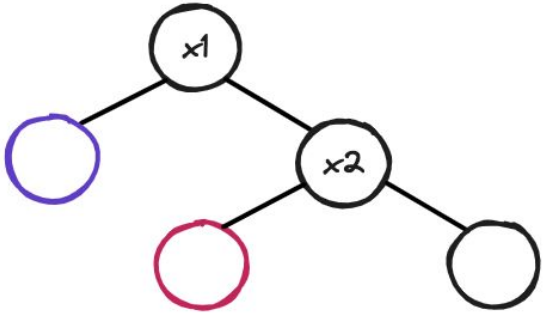

TREE
SELECT SPLIT
BASED ON IMPURITY
x2
x1
x1

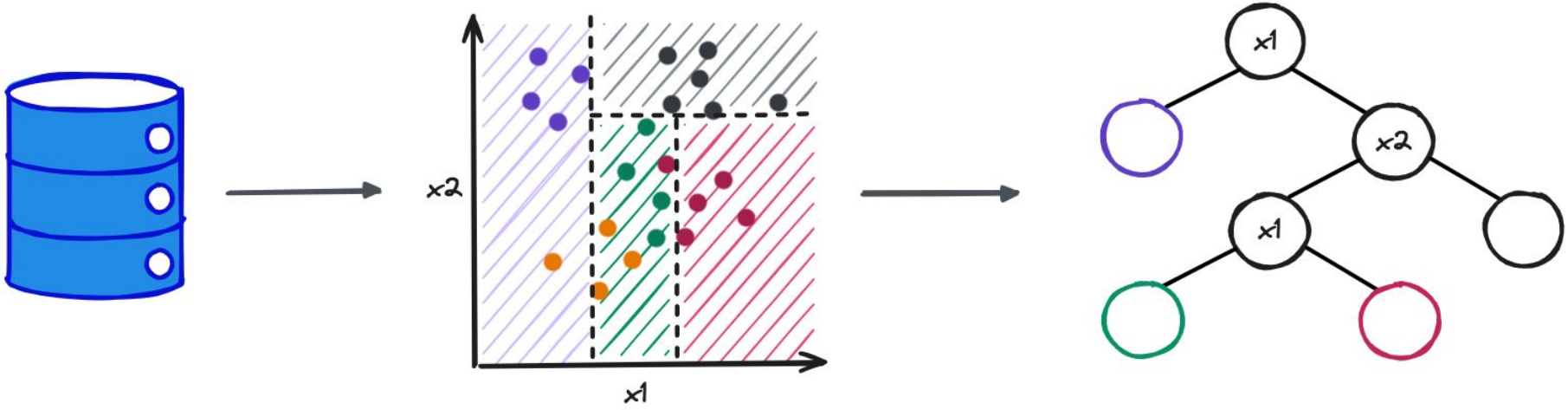


TREE

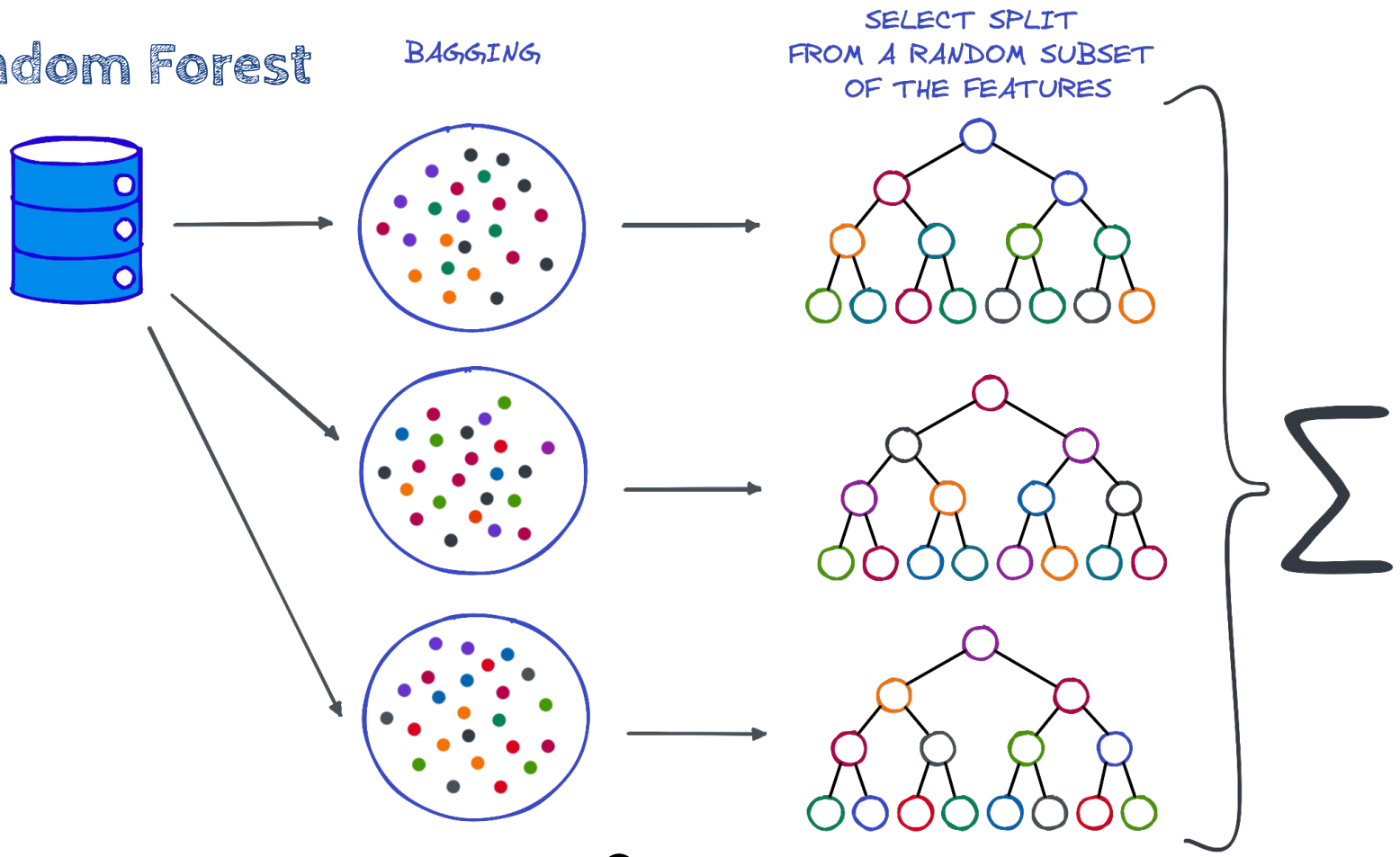


SELECT SPLIT
BASED ON IMPURITY

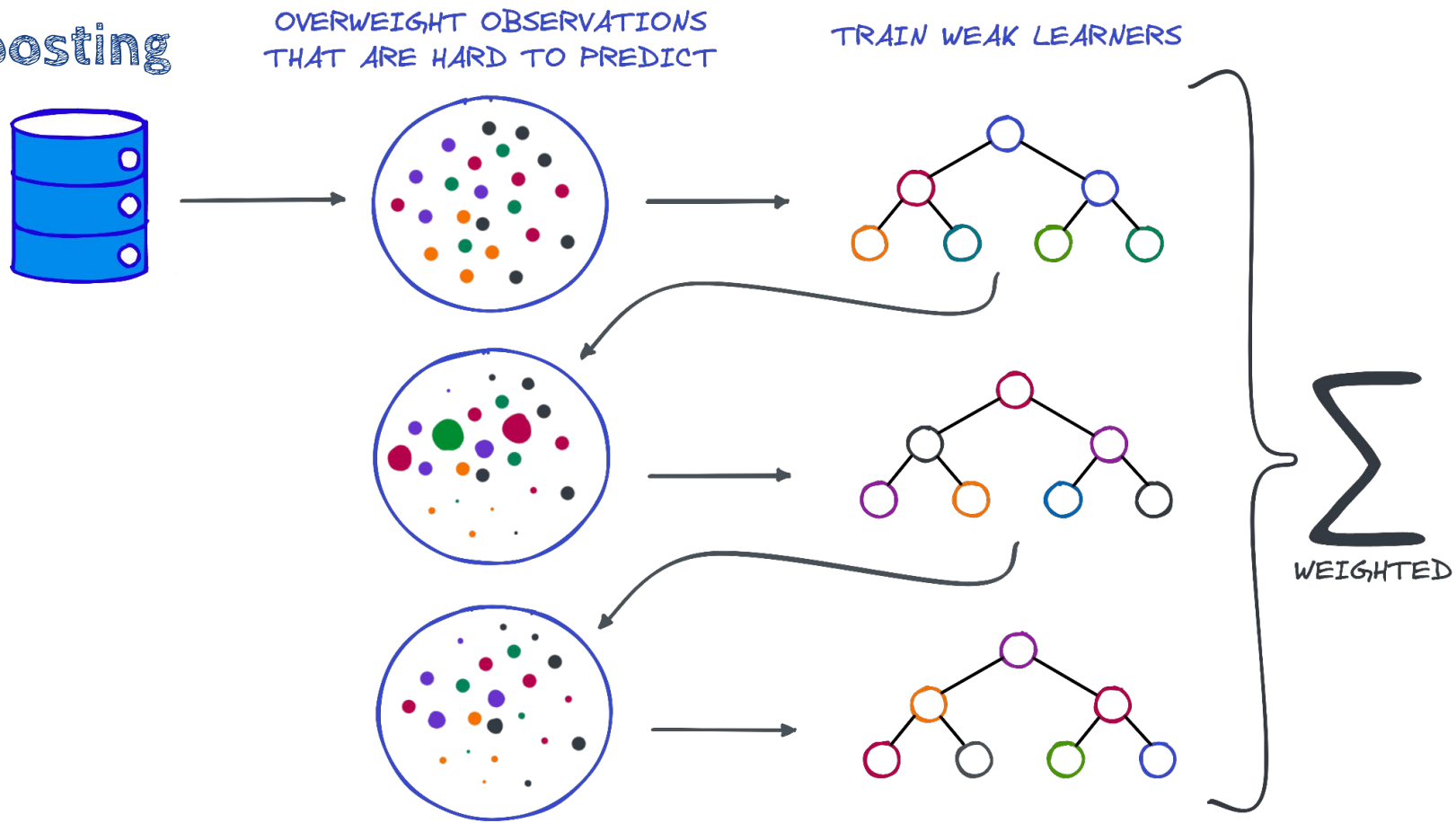



TREE
SELECT SPLIT
BASED ON IMPURITY
x2
x1
x1
x2
x1

Random Forest



Boosting



@divenyijanos

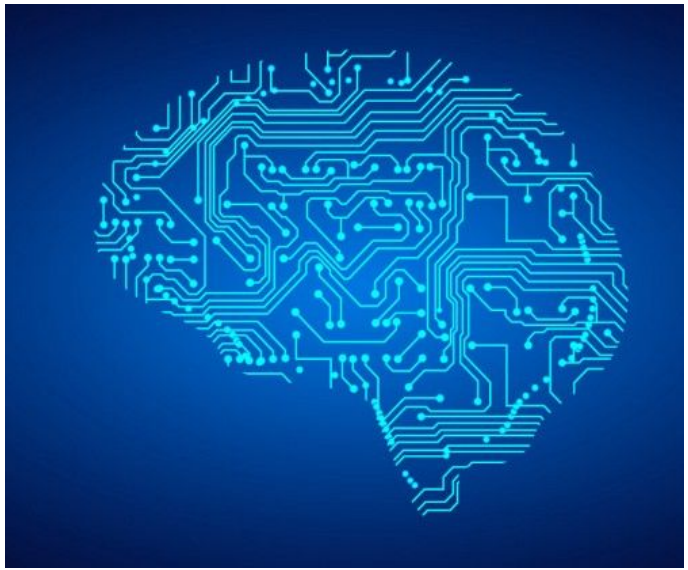
Ensemble Methods

Random Forest	AdaBoost	Gradient boosting
full trees	stumps	small trees
randomly selected features	all features	all features
bootstrap samples	whole sample	whole sample
same weight	weight by error	same weight
parallel	iterative	iterative

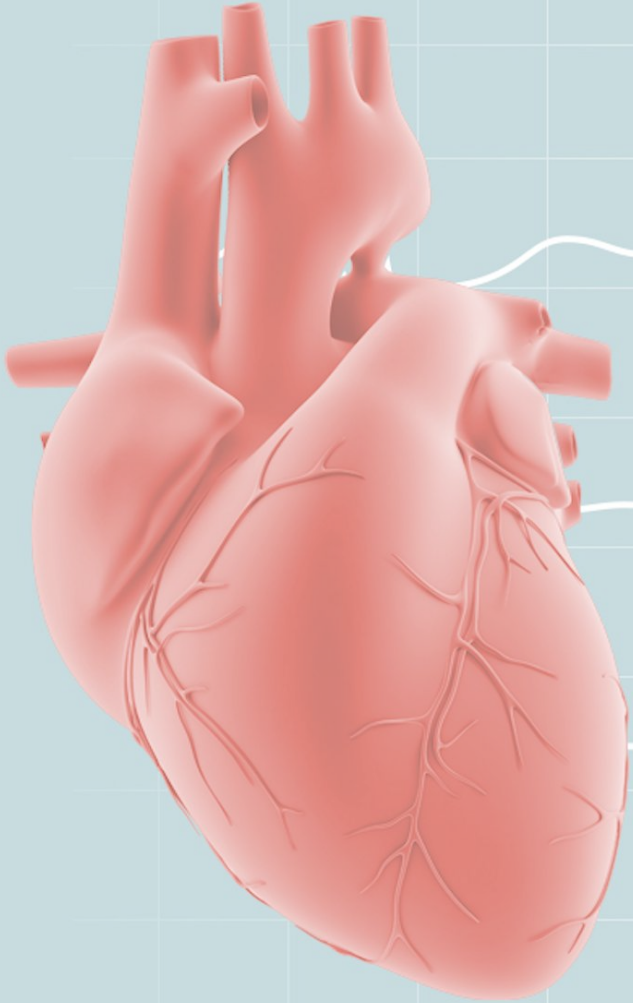

DATA
ML
~~MAGIC~~
15 5688
8 98
1257 120 36
144 784
LABELS

Loss function predicted vs real (e.g. MSE)

Optimization algorithm tweak params to minimize loss (e.g. derivation)

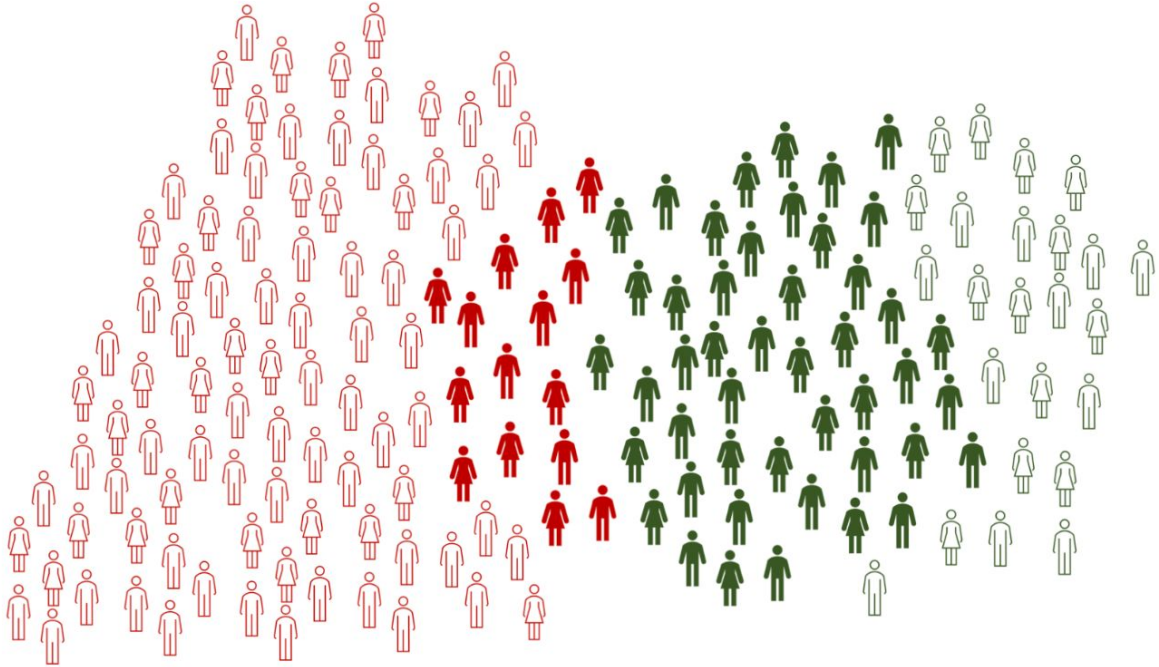


~~**Memorization**~~ vs Generalization



Heart Failure Prediction

@divenyijanos



predicted to survive

Accuracy

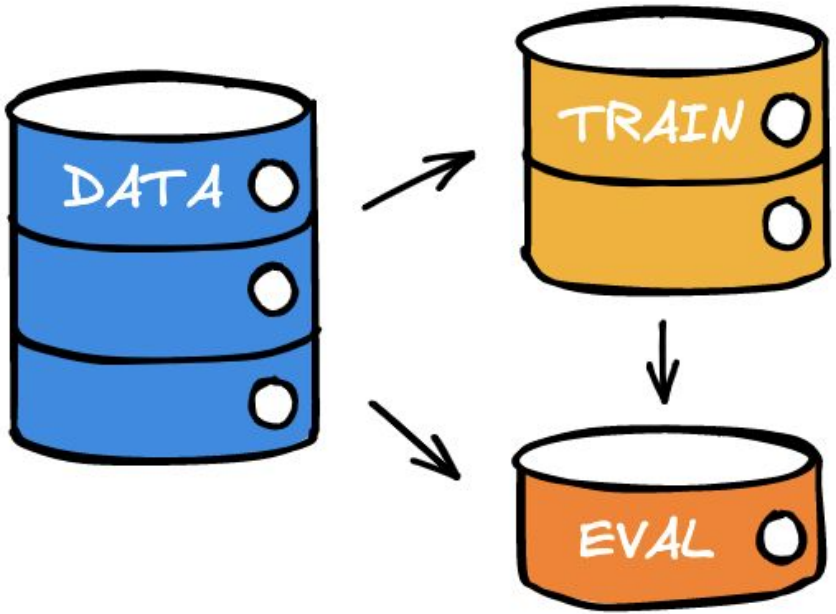
True Positive Rate

False Positive Rate

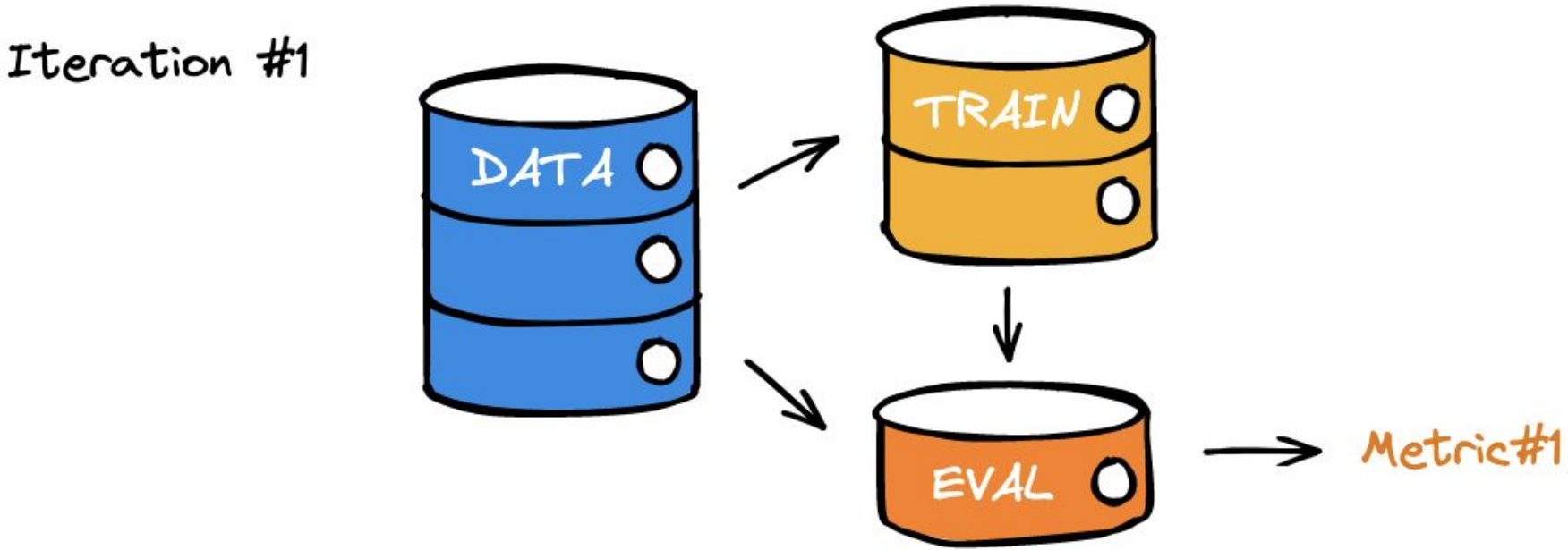
Recap: Cross Validation

Recap: Cross Validation

Iteration #1

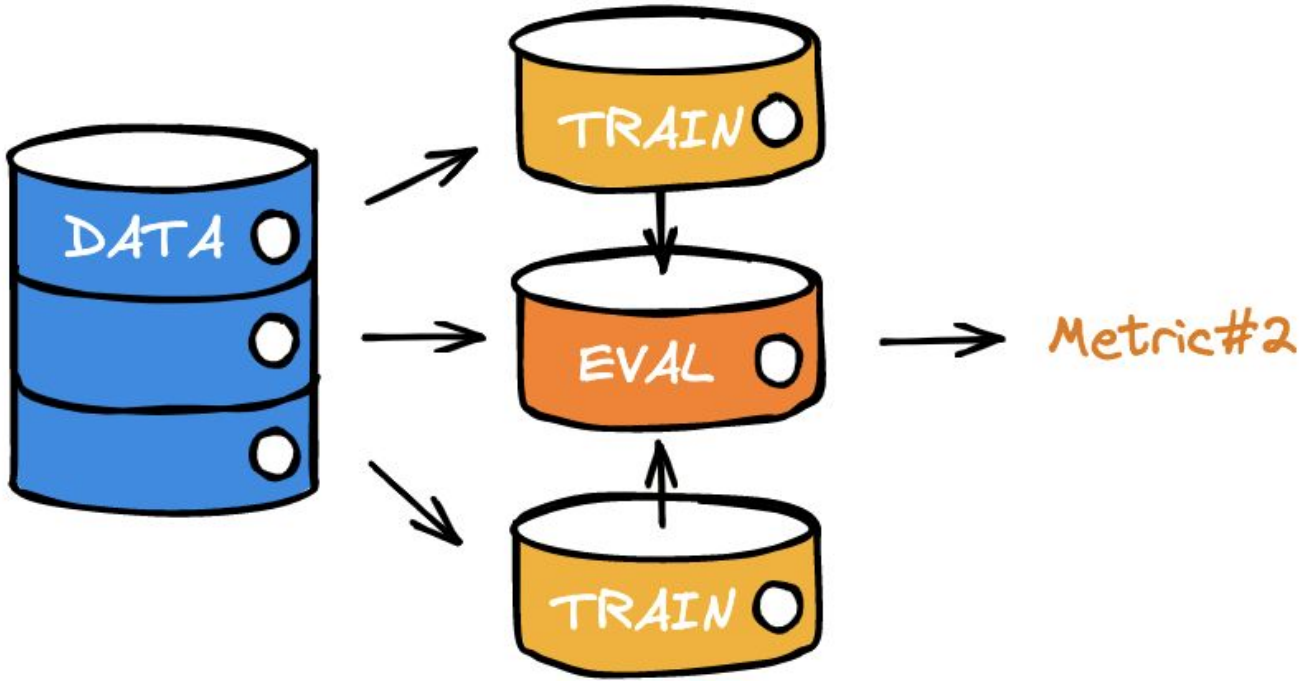


Recap: Cross Validation

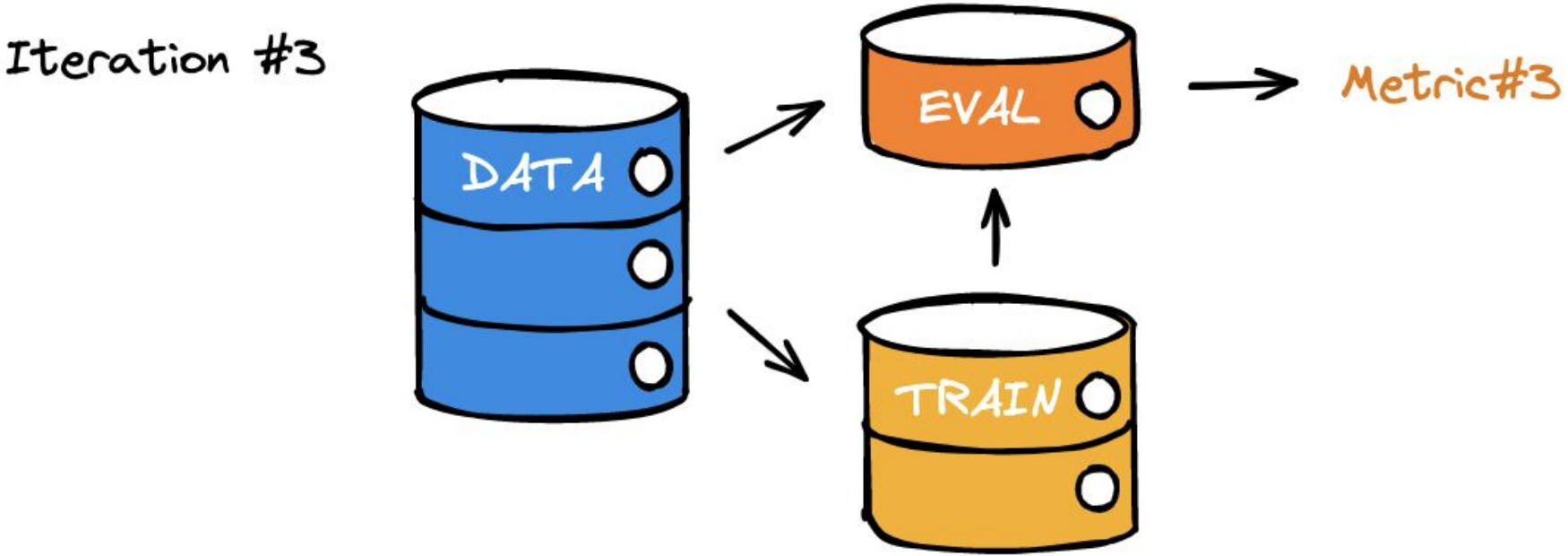


Recap: Cross Validation

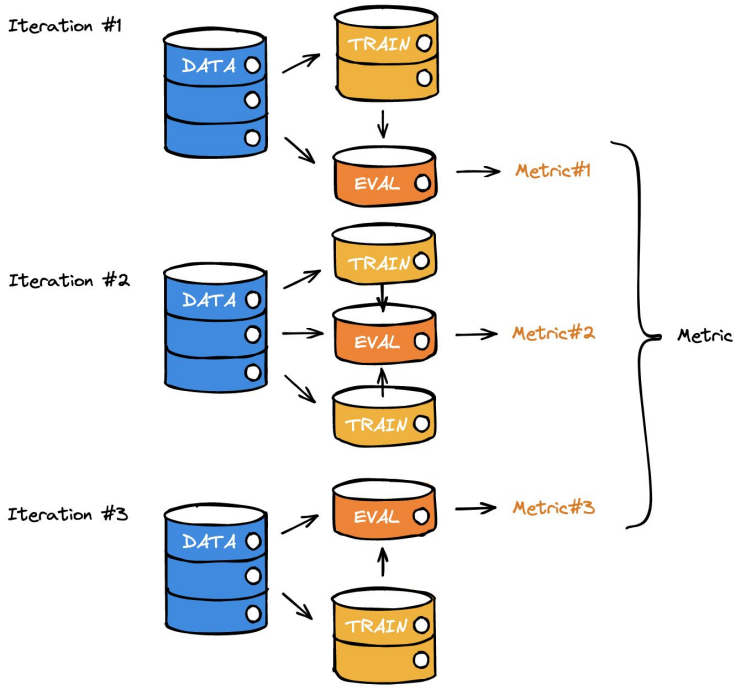
Iteration #2



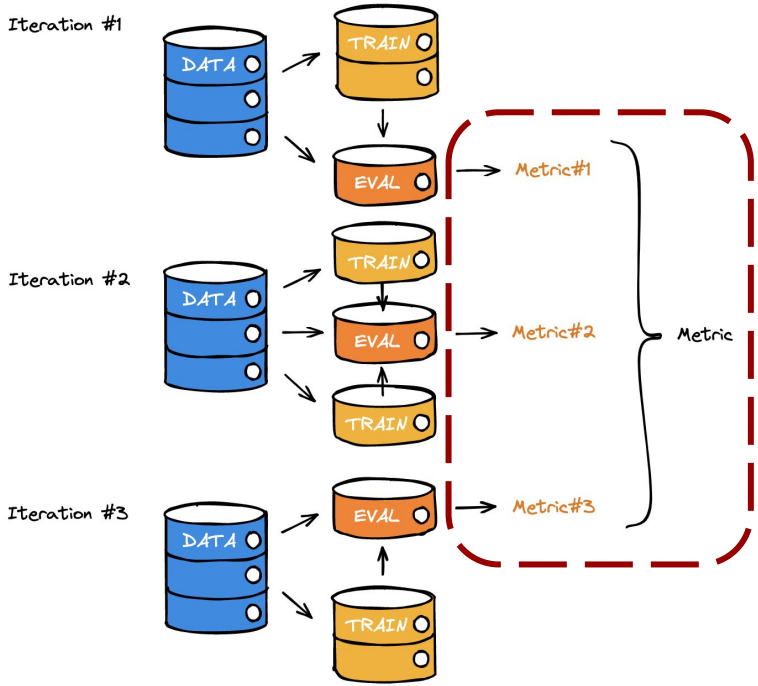
Recap: Cross Validation

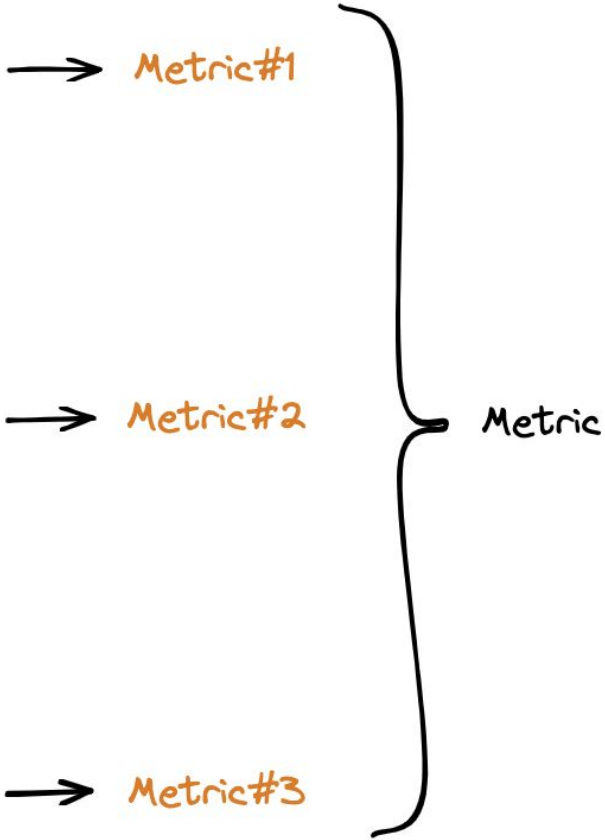


Recap: Cross Validation



Recap: Cross Validation





Recommended Materials

Video:

- Grant Sanderson (3Blue1Brown): Neural Networks (4 videos)