

Machine Learning

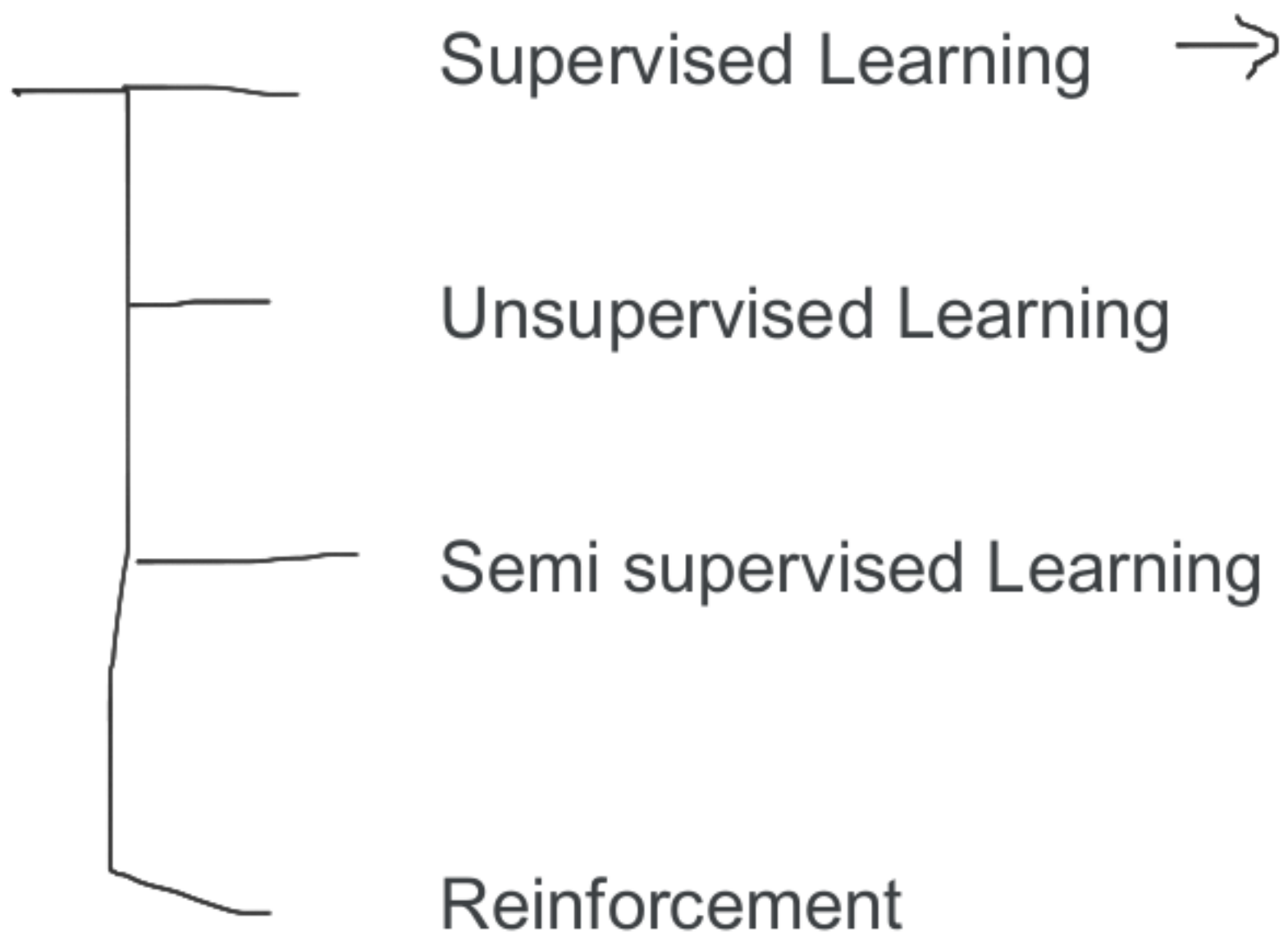
Problem Identification -> Data Collection -> Data processing -> EDA -> Feature engineering

ML

gpa	label
3.5	Pass
1.2	Fail
2.8	Fail
3.8	Pass
4.0	pass
3.0	Pass

gpa	label
3.5	Pass
1.2	Fail
2.8	Fail
3.8	Pass
4.0	pass
3.0	Pass

ML



Supervised Learning



Robot (ML Model)



image



Cat image



Label

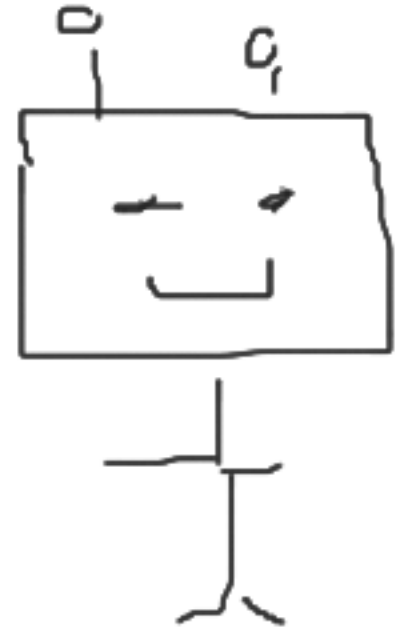
Result label

Training Time



F1 - F9
- cat
- dog

Supervised Learning



Robot (ML Model)

image



Cat image

Label

Result label

Training Time

Features

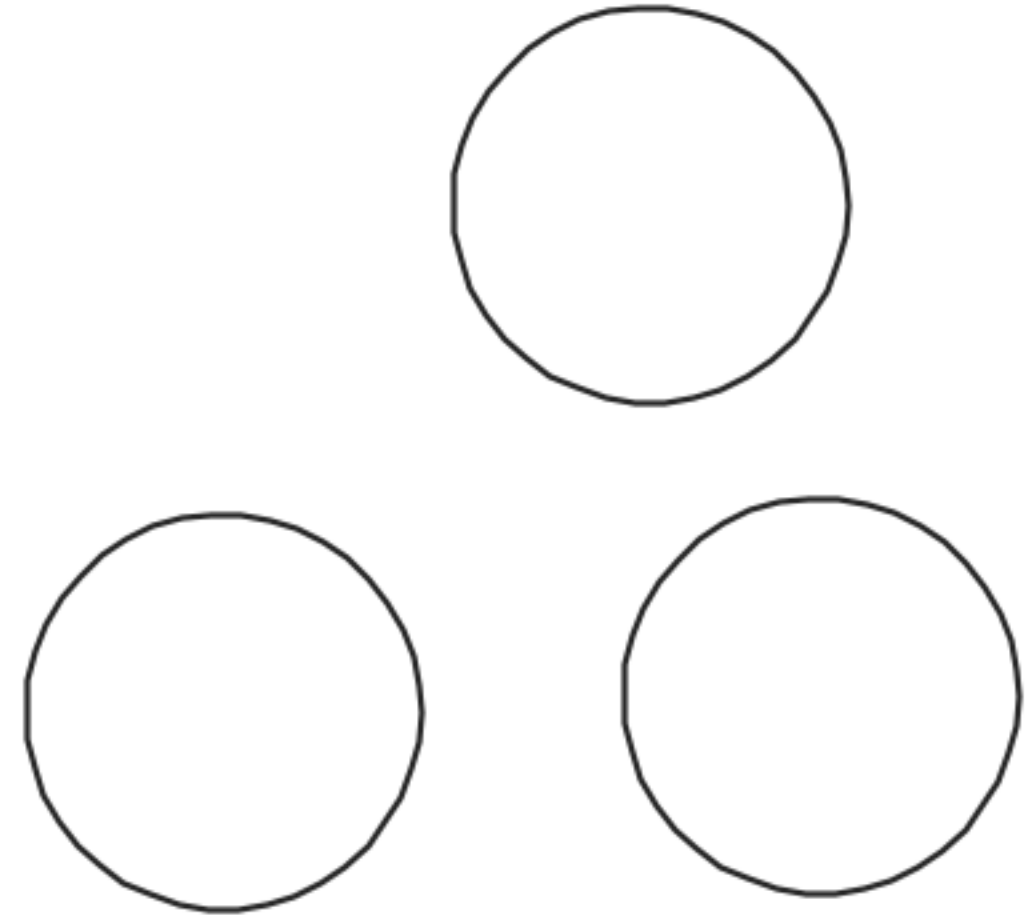


cat

Un supervised learning

No labeled data

you will only have features
-> result will be in form of Clusters



Features
+
Labels



MODEL

SL

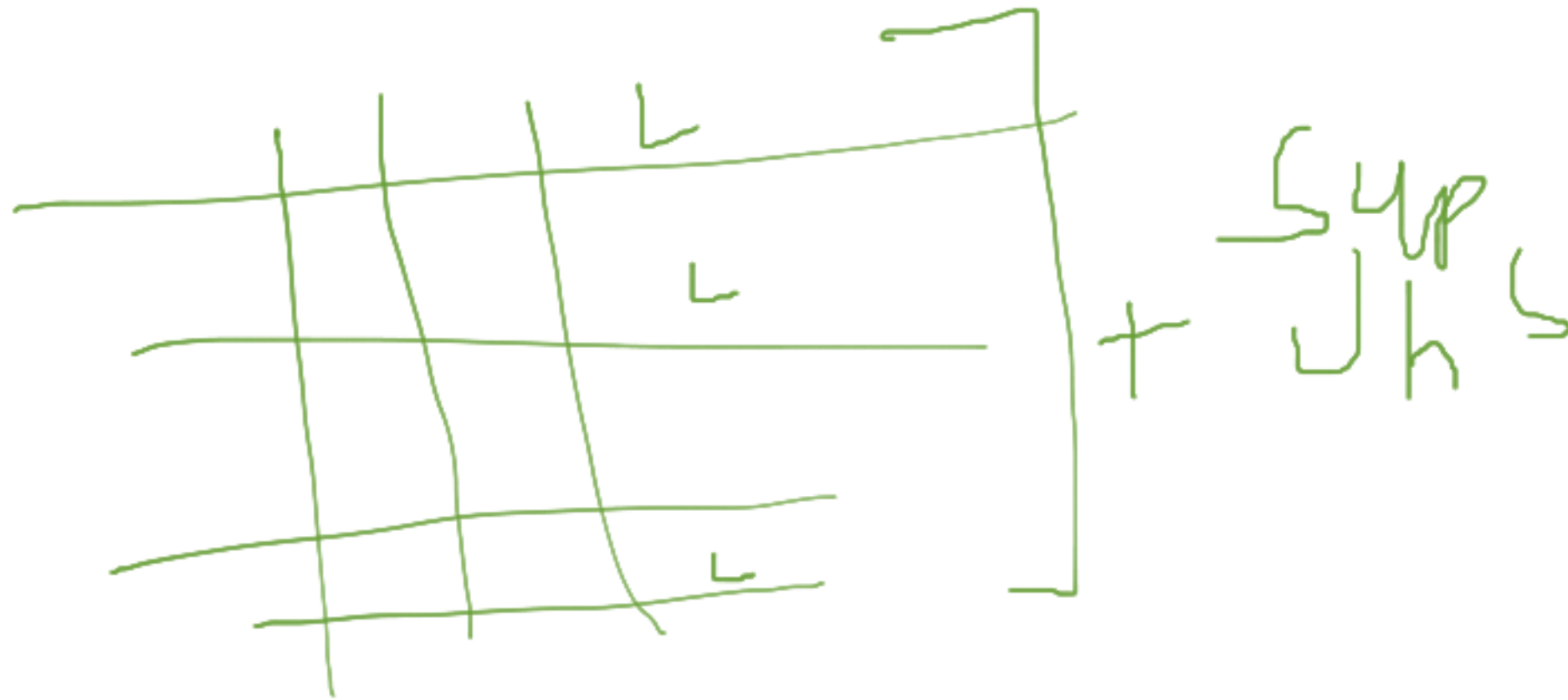


F					C
					C
					D
					L



UNS





Semi
supervised

gpa	label
3.5	Pass
1.2	Fail
2.8	Fail
3.8	Pass
4.0	pass
3.0	Pass

gpa	label
3.5	Pass
1.2	Fail
2.8	Fail
3.8	Pass
4.0	pass
3.0	Pass

F T L
 F T L

supervised
learning

gpa

3.5
1.2
2.8
3.8
4.0
3.0

F =
 F =

Un-Supervised
learning

gpa	label
3.5	Pass
1.2	Fail
2.8	
3.8	Pass
4.0	
3.0	

→

Semi
Supervised
Learning

Re inforcement Learning

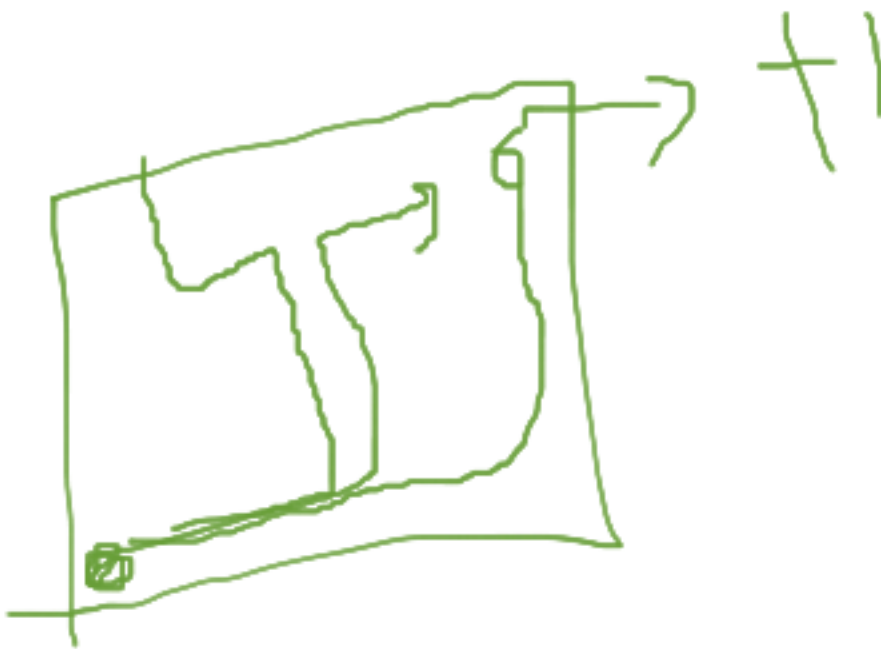
→ AI



Test and Trial

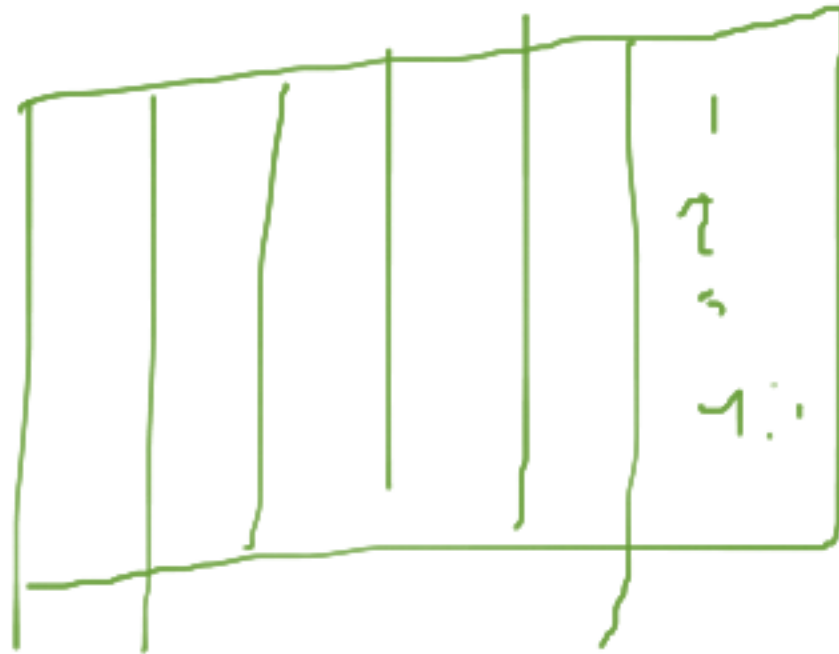
it works on reward
and penalty

end ↑ +1



Supervised learning

Prediction /
regression



					1
					2
					3
					4.1

not a
category

continuous
values
(numeric)
price,

repeated
fixed values

categories

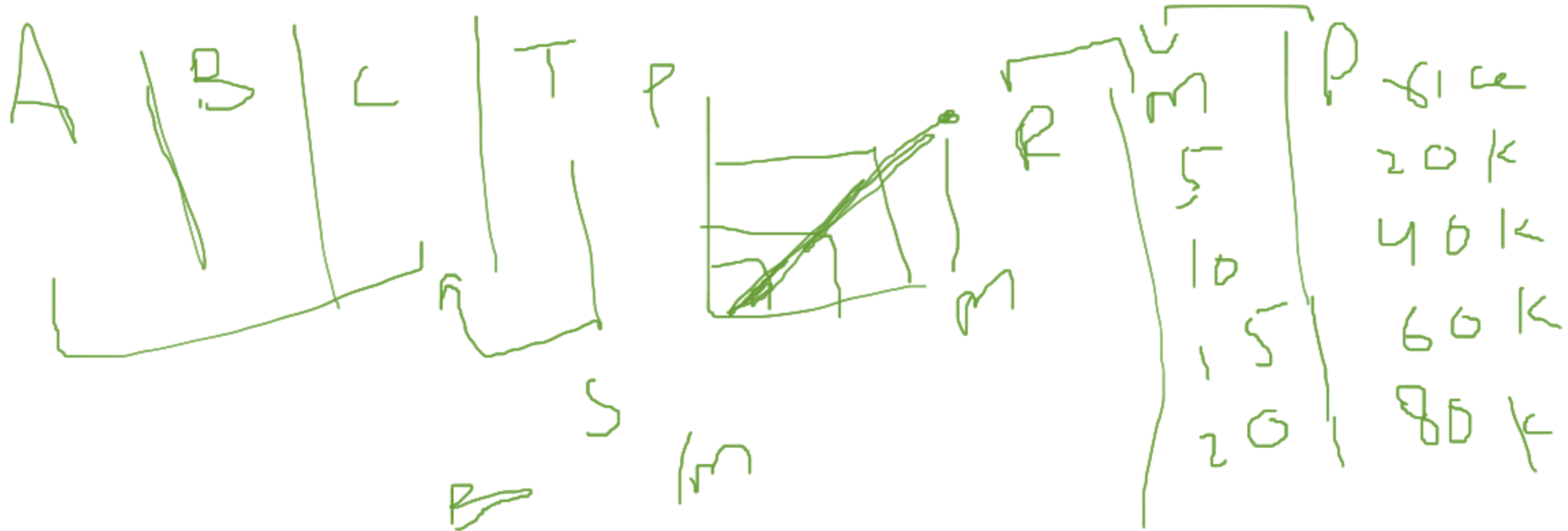
Classification



					1
					3
					2
					1

category

Binomial



Relation ?

if x has one column \Rightarrow Simple linear regression

if x has more than one column \Rightarrow Multiple linear regression

Data



Identify the target columns



✓ F L

Split data into test and train

Data => training + testing samples

80 - 20 ✓

70 - 30

90 - 10

Training	80	test	20%
----------	----	------	-----

x=80%
y=80%

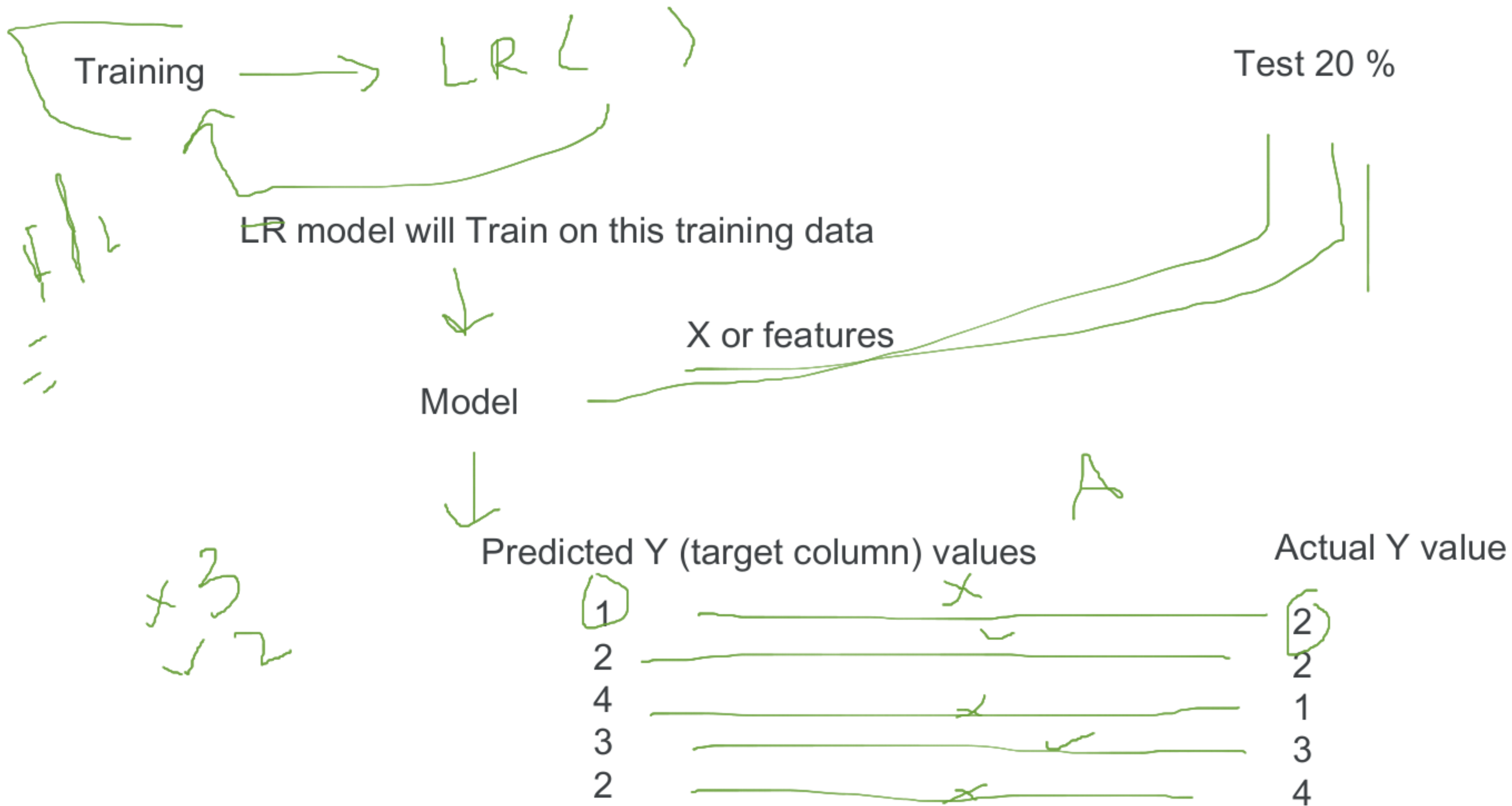


x=20%
y=20%

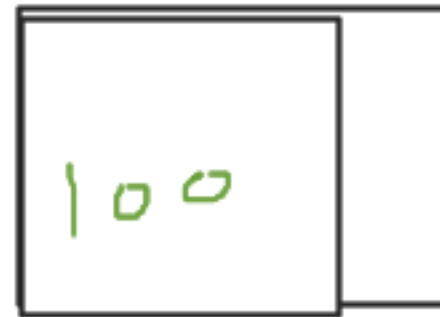


Test





Data



identify target column



Split into train and test

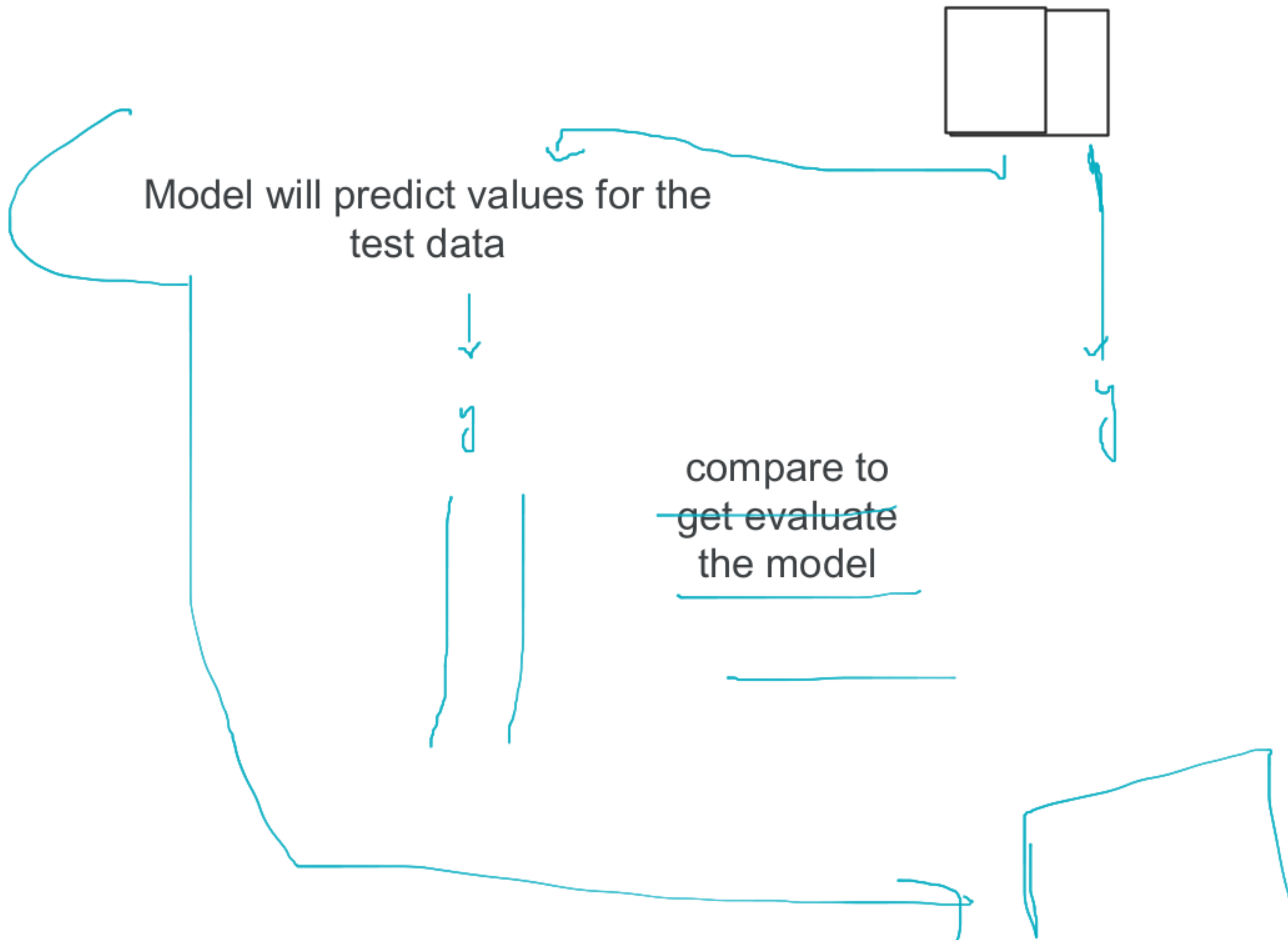


Train the model on above 80%



Model will predict values for the
test data





Age
18, 19, 17, 20, 21 — test
train

A diagram illustrating data partitioning for a variable named 'Age'. The values '18, 19, 17, 20, 21' are listed. A bracket underneath the first three values ('18, 19, 17') is labeled 'train'. A circle around the last two values ('20, 21') is labeled 'test'.

17-19 teen
ages

Data 100 %
train:80%
test:20%

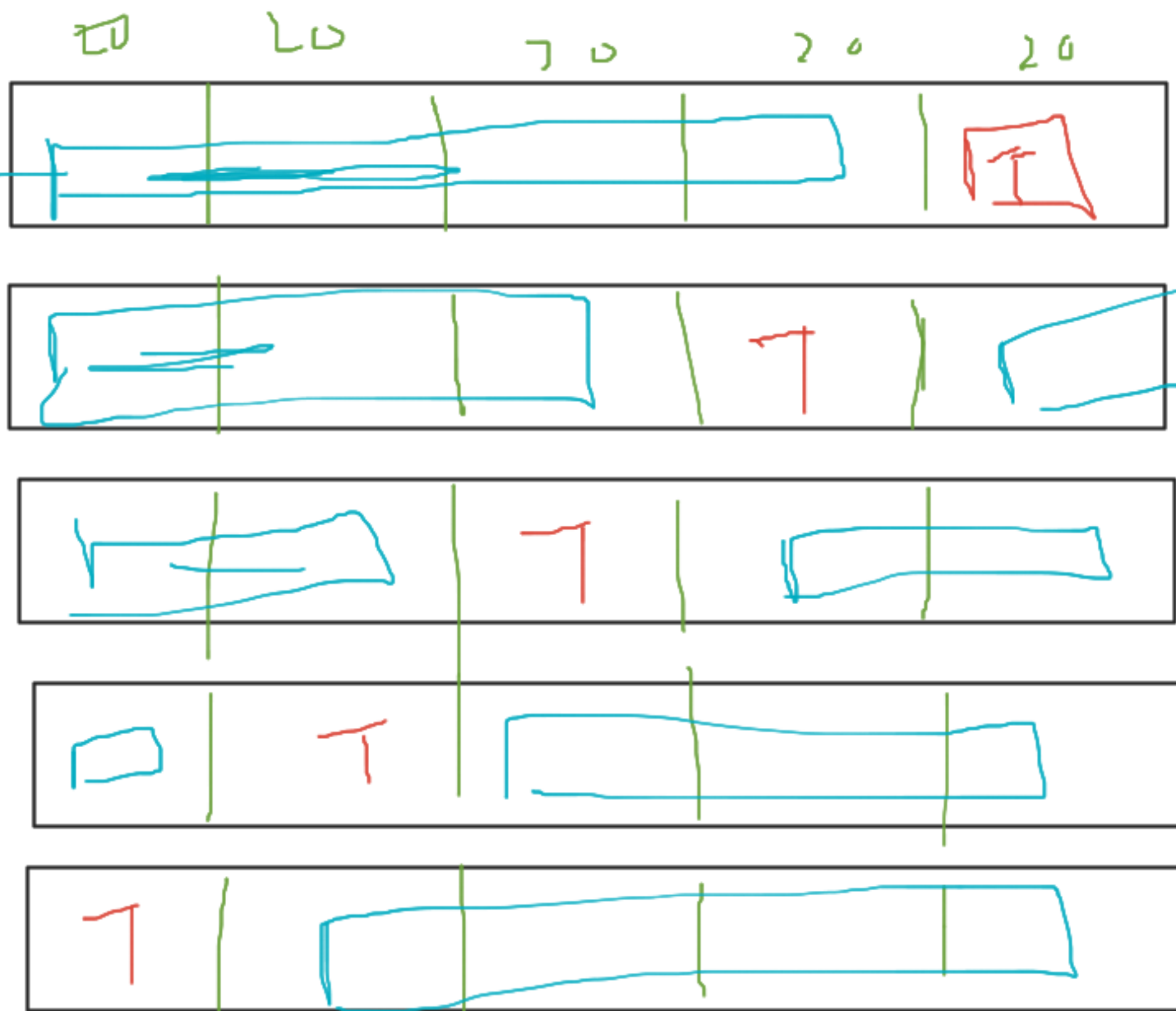
Cross validation

→ 13

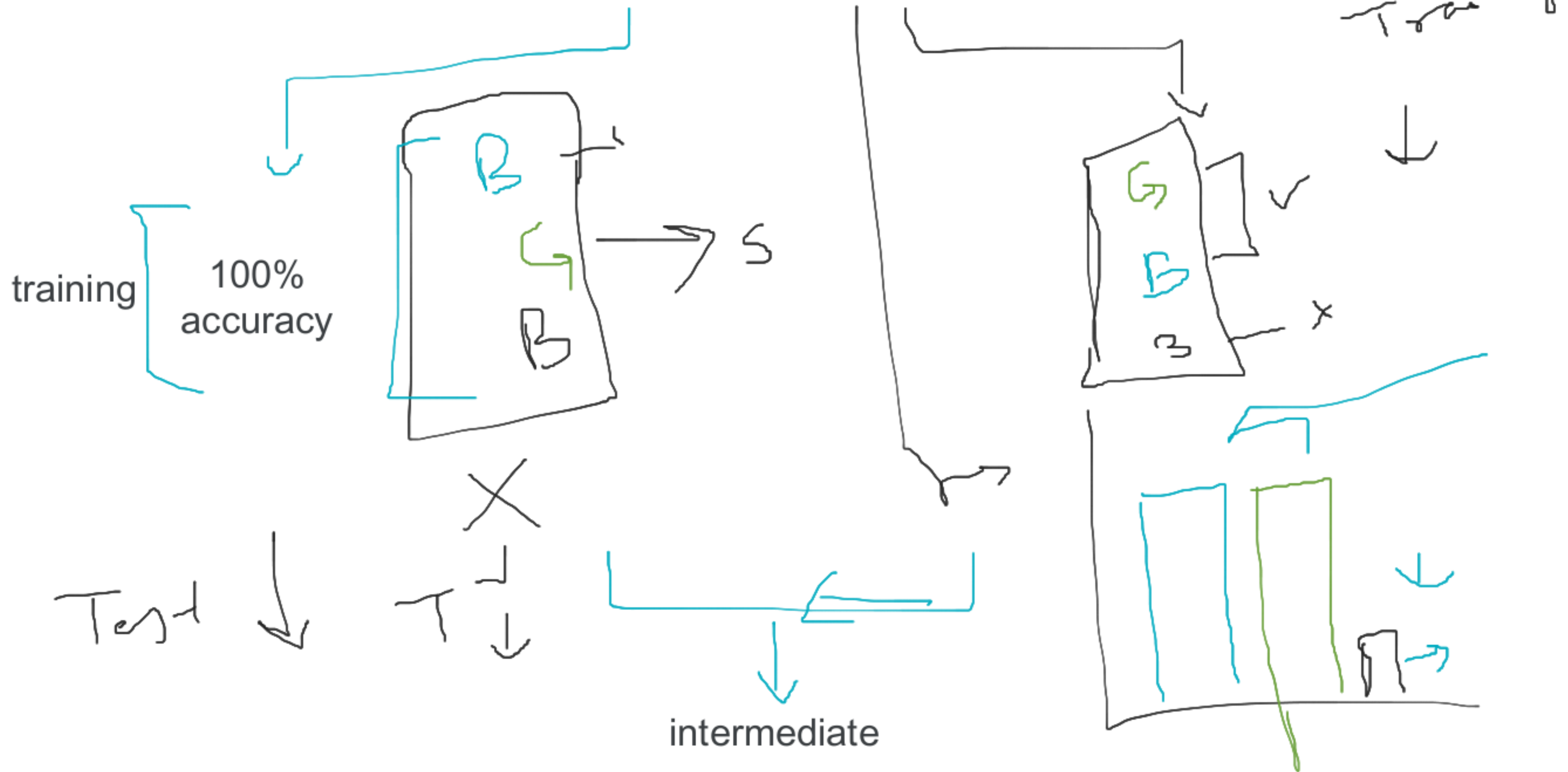
fold = 5

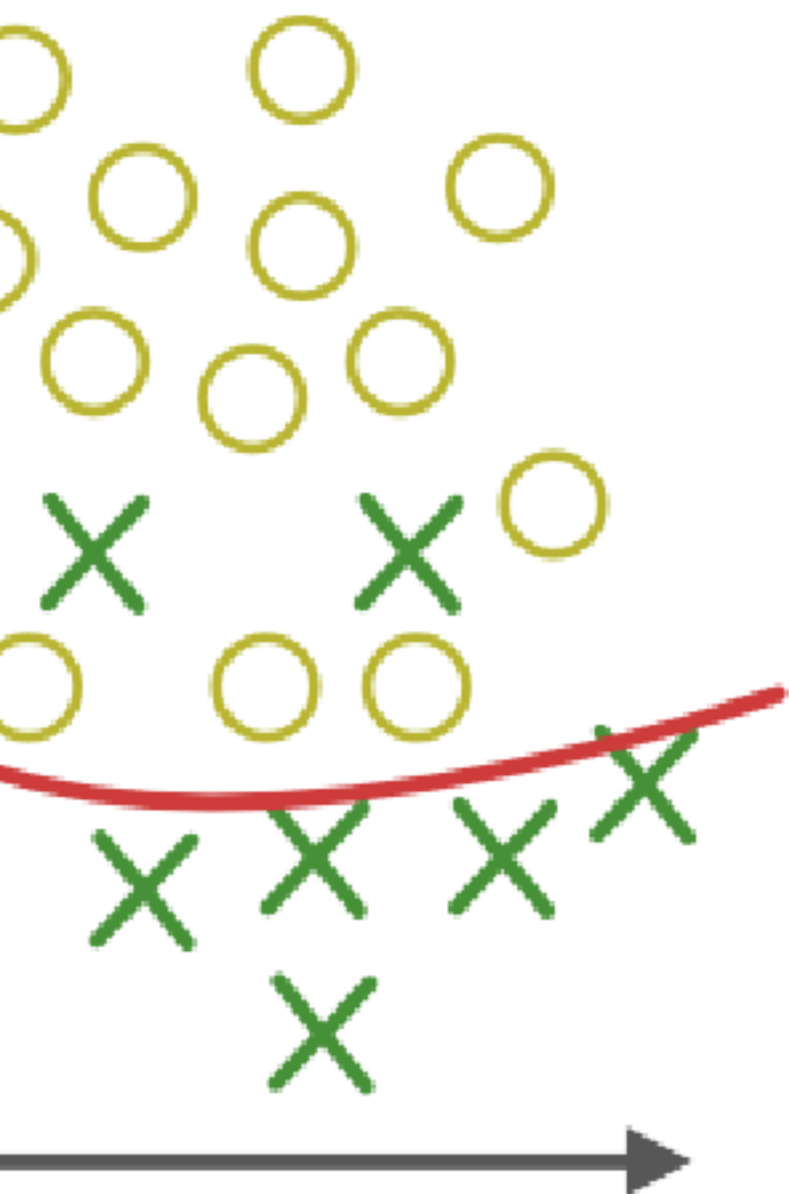
$100 \% / 5$
 $= 20\%$

training

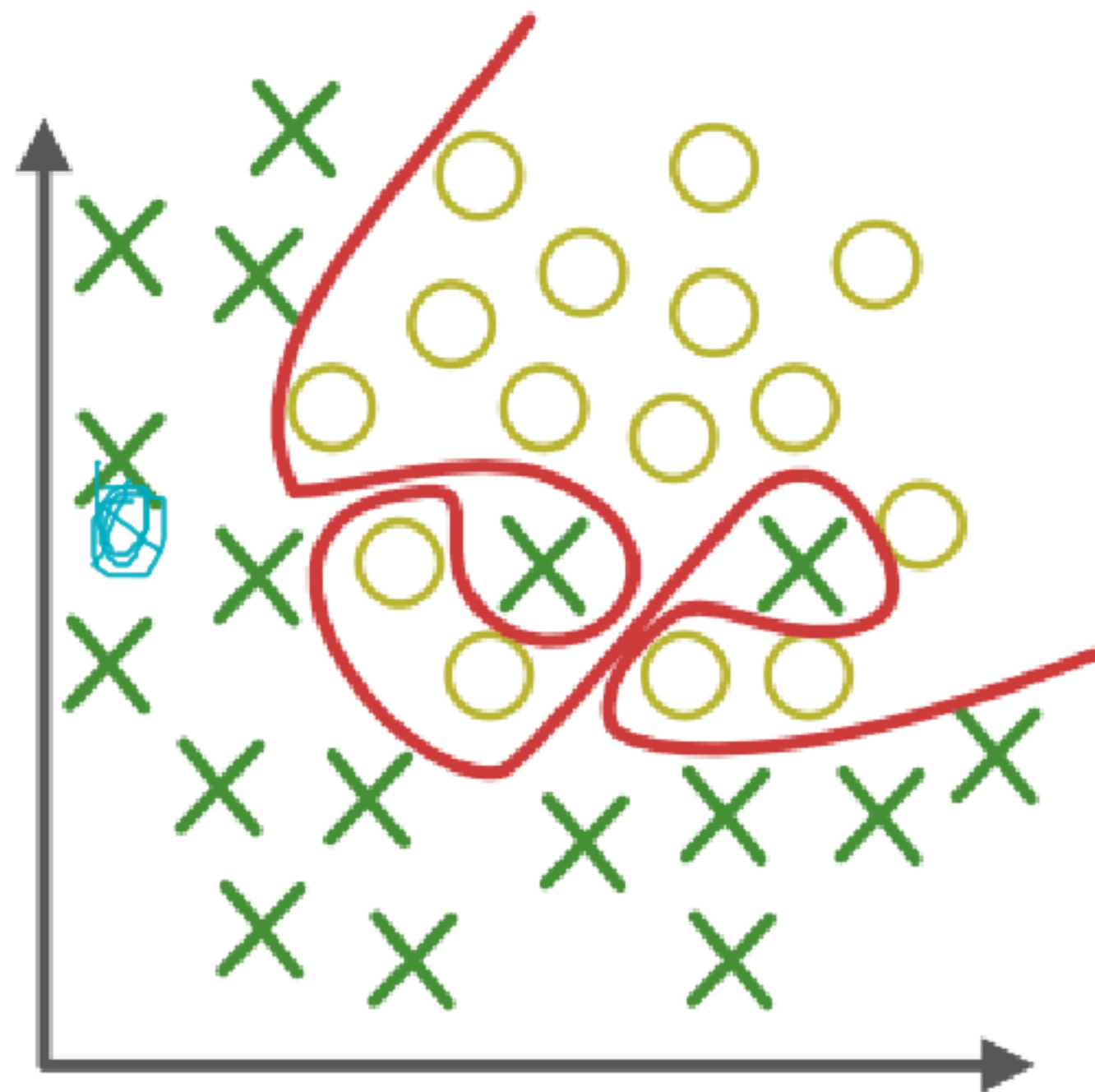


Overfitting vs underfitting





Linear fitting



Over-fitting

(forcefitting--too
good to be true)

