Introduction to machine learning

Andrea Massari - 07/26/2017

Vocabulary

(by the end of the lecture you'll know what these words mean)

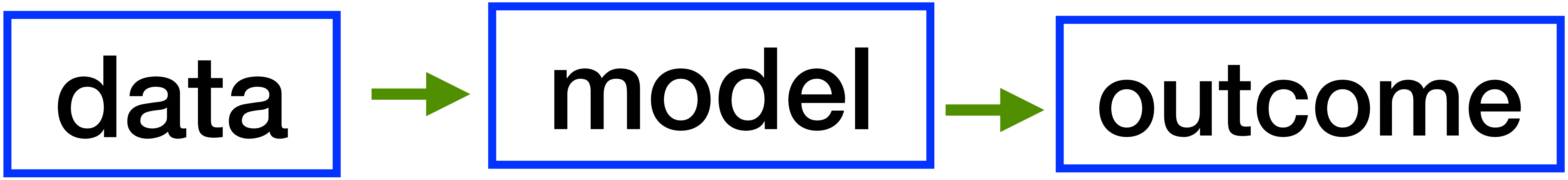
- (cross/k-fold)validation
- hyperparameter
- algorithm
- parameter
- classification
- regression
- likelihood
- feature
- loss/cost function
- model
- performance
- testing
- overfitting
- underfitting
- class
- label
- …

Why Machine Learning?

1. teach a machine to do a human task
   e.g. everyone can distinguish a cat from a dog, can we teach a computer to do it?

2. teach a machine to do “super”-human tasks
   e.g. distinguish two people from the way they walk

ML project

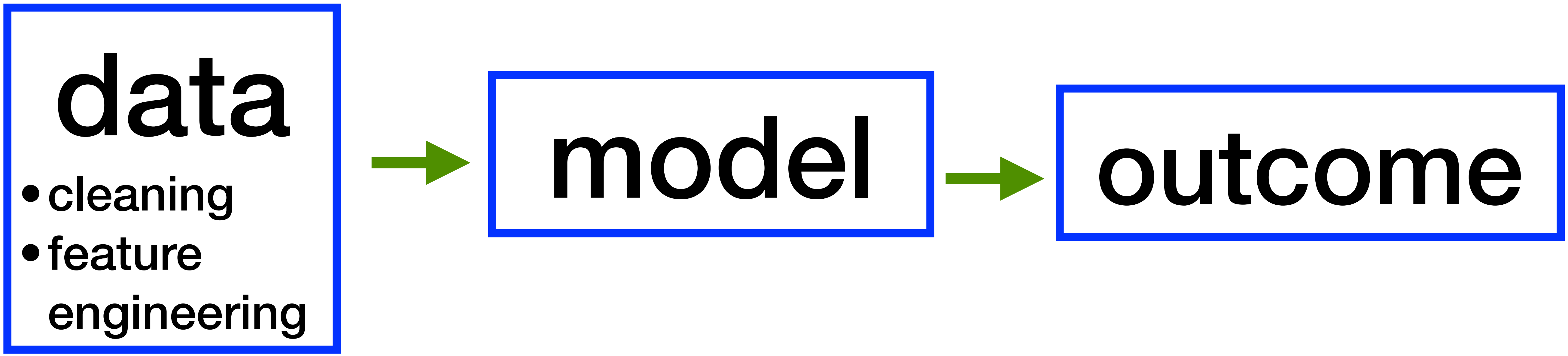


data

model

outcome

ML project



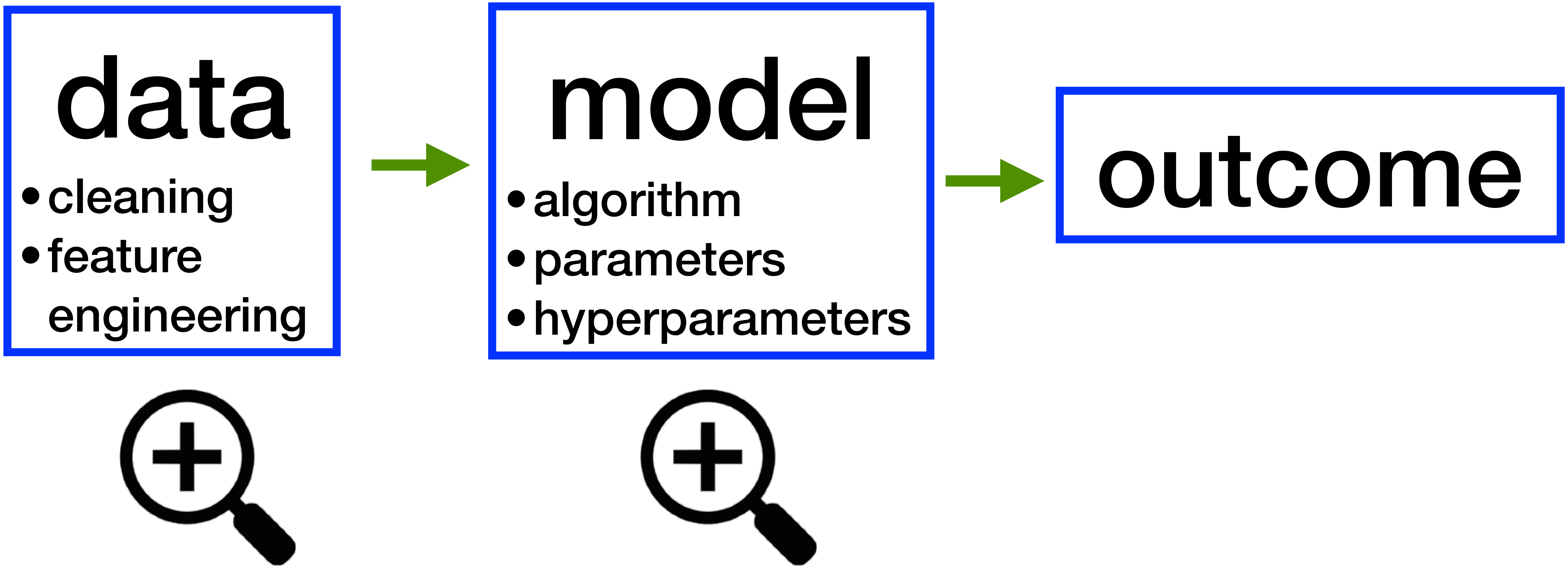
data

- cleaning
- feature engineering

model

outcome

ML project



data

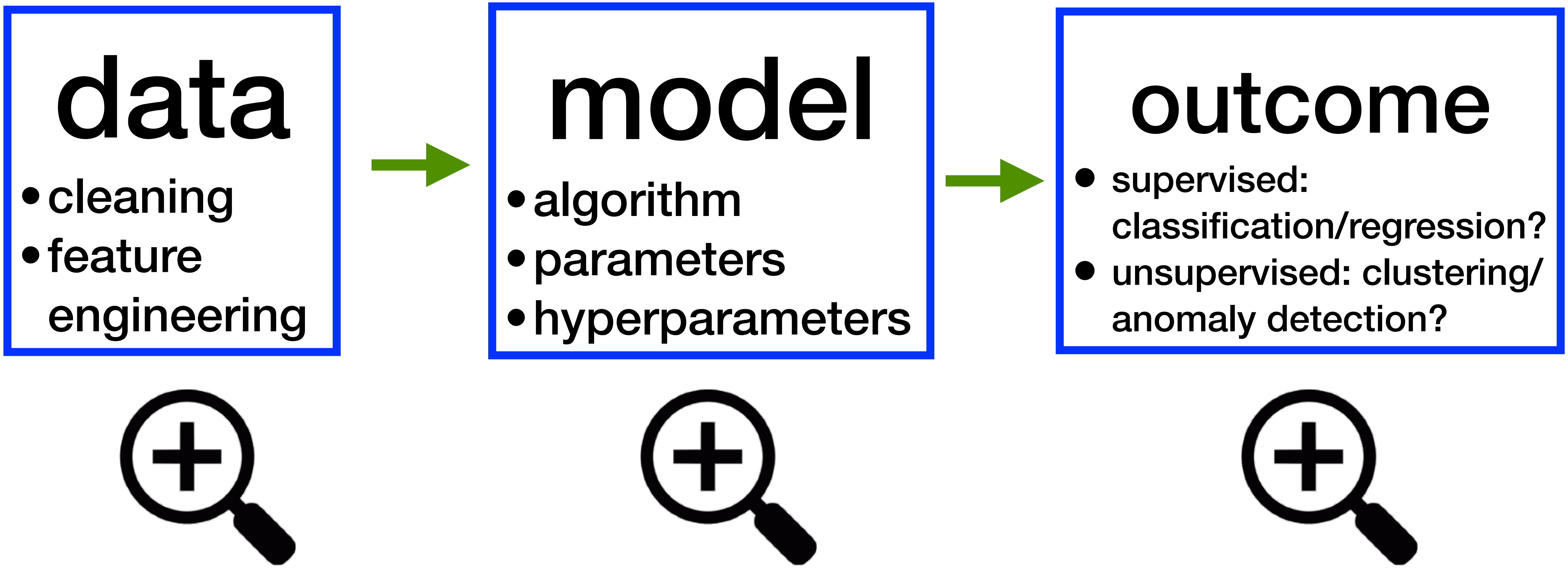
- cleaning
- feature engineering

model

- algorithm
- parameters
- hyperparameters

outcome

ML project



data

- cleaning
- feature engineering

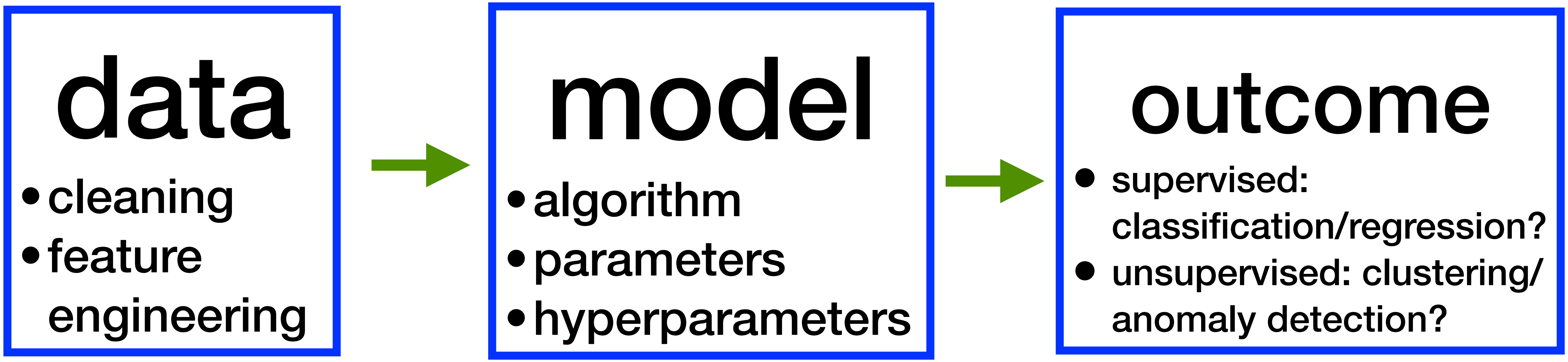
model

- algorithm
- parameters
- hyperparameters

outcome

- supervised: classification/regression?
- unsupervised: clustering/ anomaly detection?

What are we “learning”?



data

- cleaning
- feature engineering

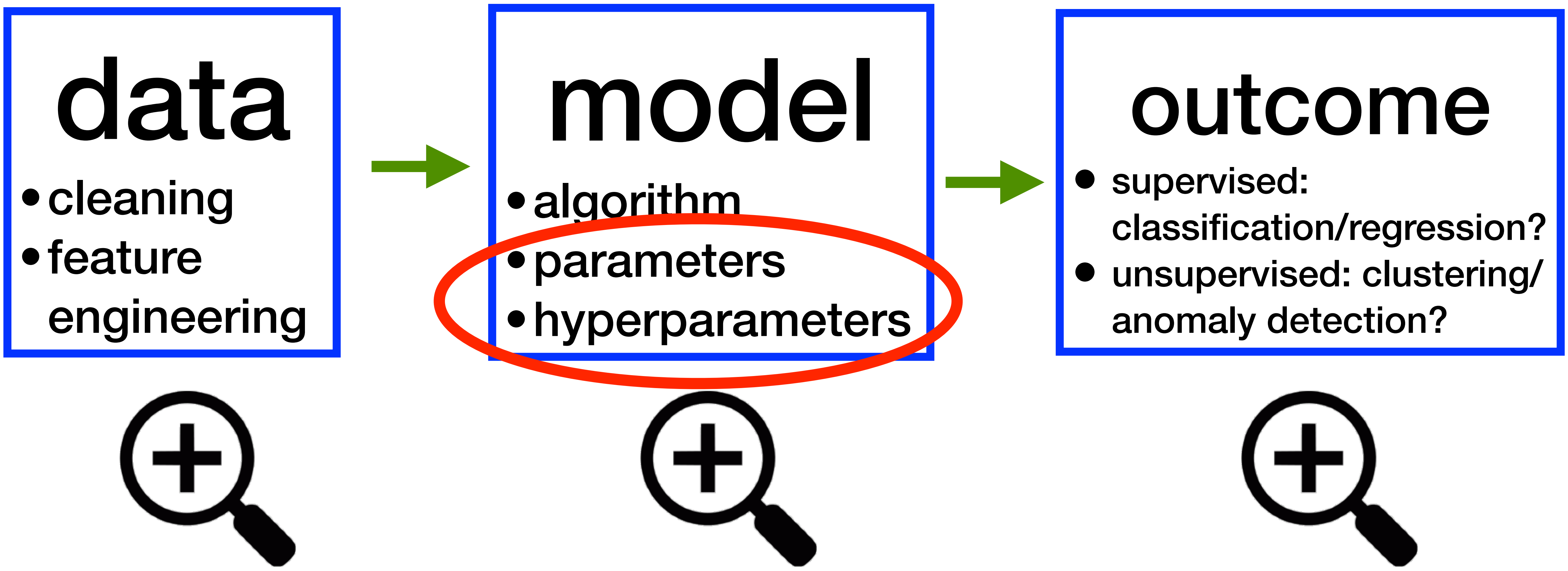
model

- algorithm
- parameters
- hyperparameters

outcome

- supervised: classification/regression?
- unsupervised: clustering/anomaly detection?

What are we “learning”?



data

- cleaning
- feature engineering

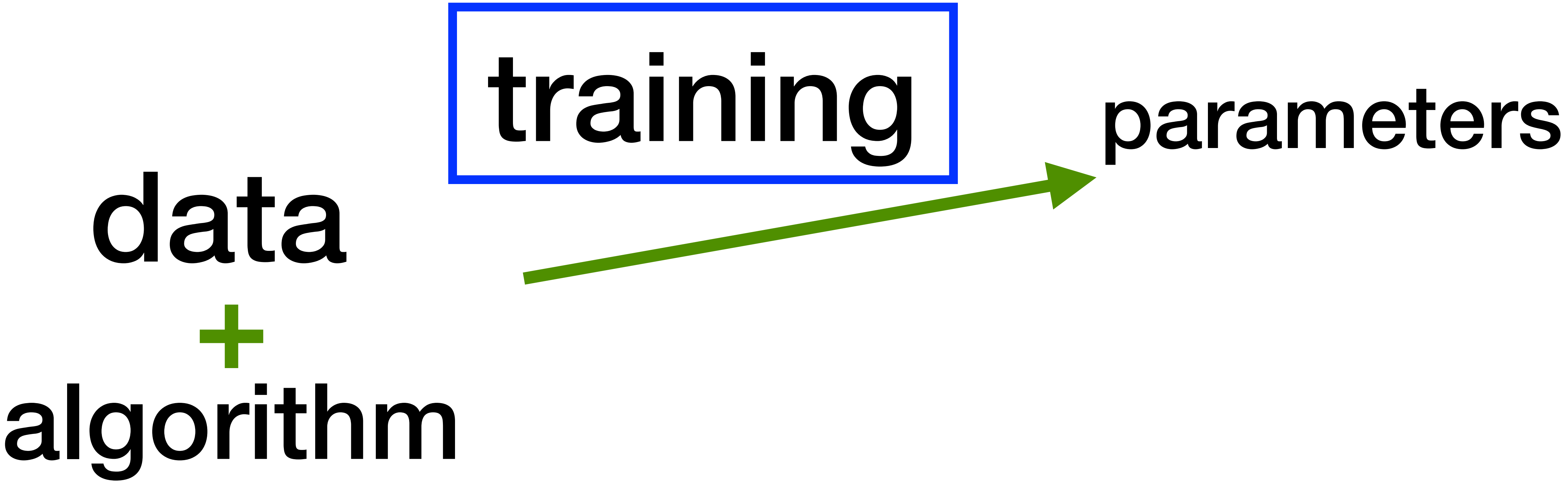
model

- algorithm
- parameters
- hyperparameters

outcome

- supervised: classification/regression?
- unsupervised: clustering/ anomaly detection?

How do we “learn” it?



training

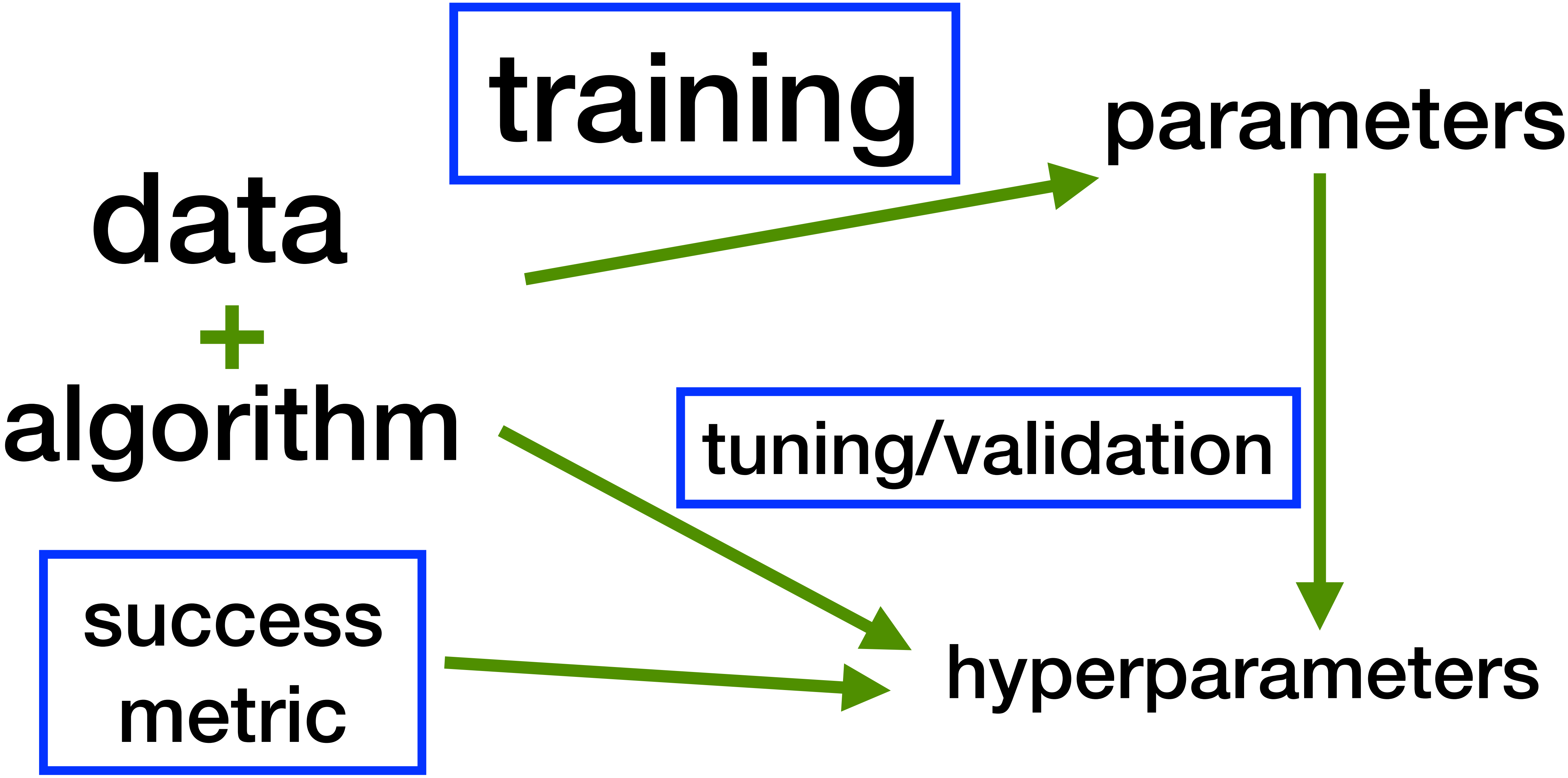
data

+

algorithm

parameters

How do we “learn” it?



training

parameters

data

+

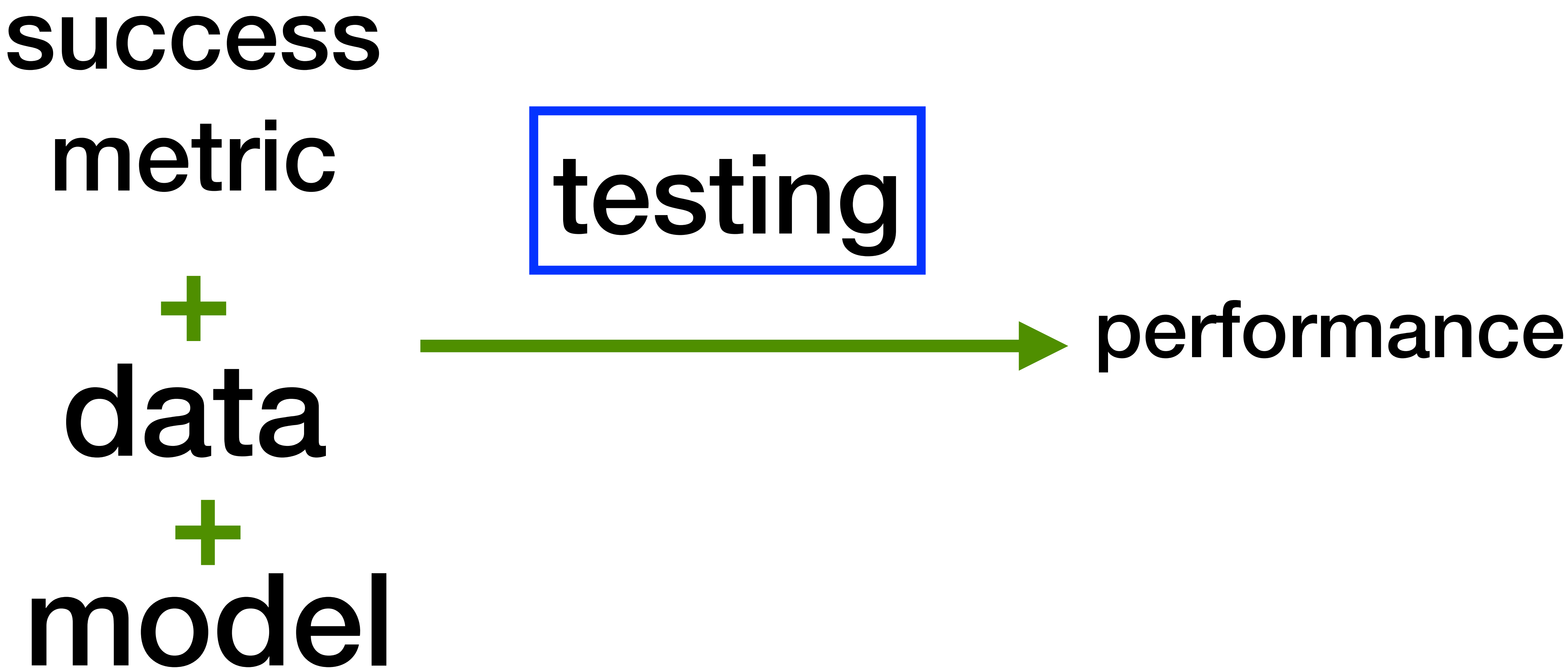
algorithm

tuning/validation

success metric

hyperparameters

Are we happy with it?



success metric

+

data

+

model

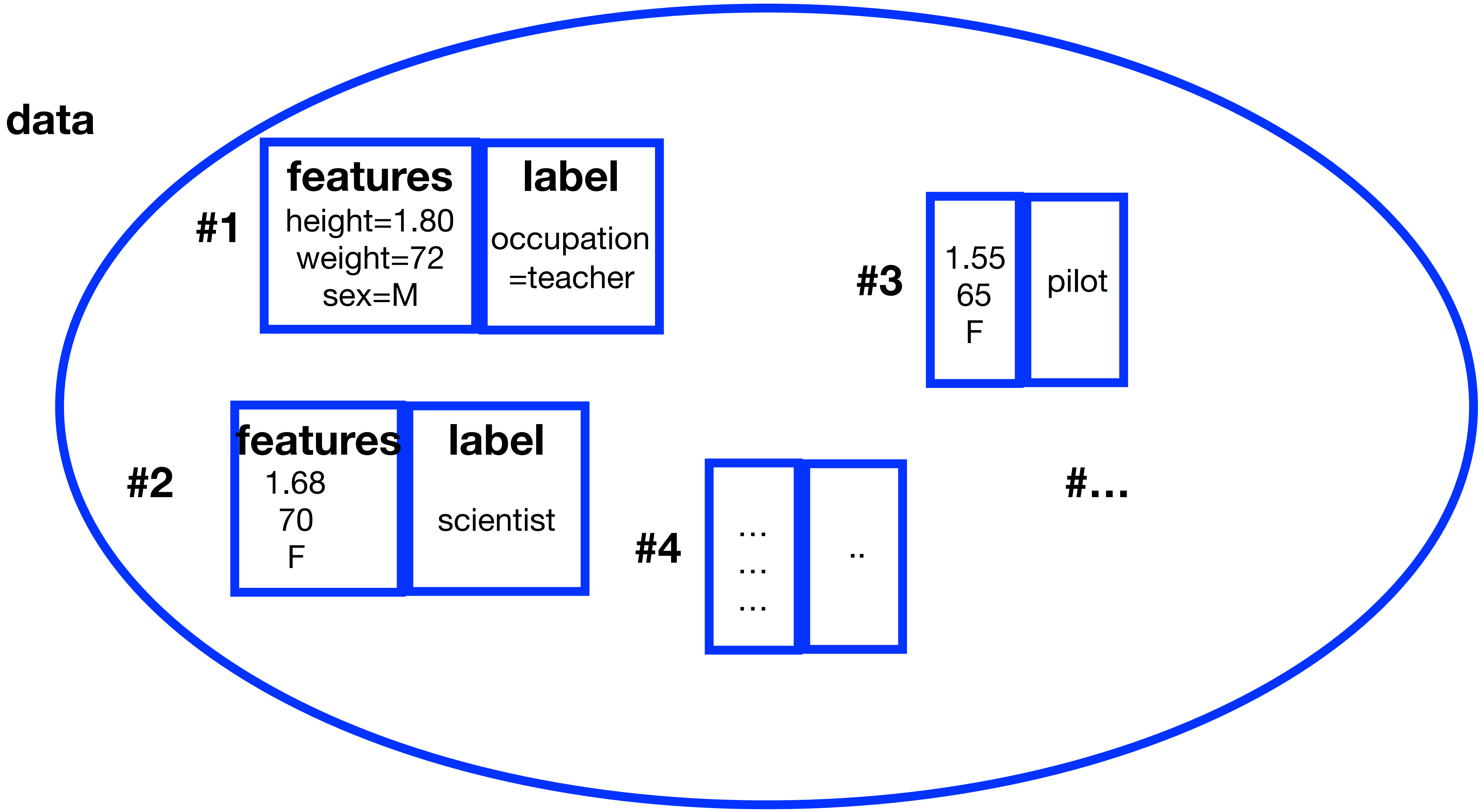
testing

performance

Supervised vs. Semi-supervised vs. Unsupervised

Supervised vs. Semi-supervised vs. Unsupervised

Suppose you want to predict occupation based on physical features (which is silly ;))



data

#1
features
height=1.80
weight=72
sex=M
label
occupation
=teacher

#2
features
1.68
70
F
label
scientist

#3
1.55
65
F
pilot

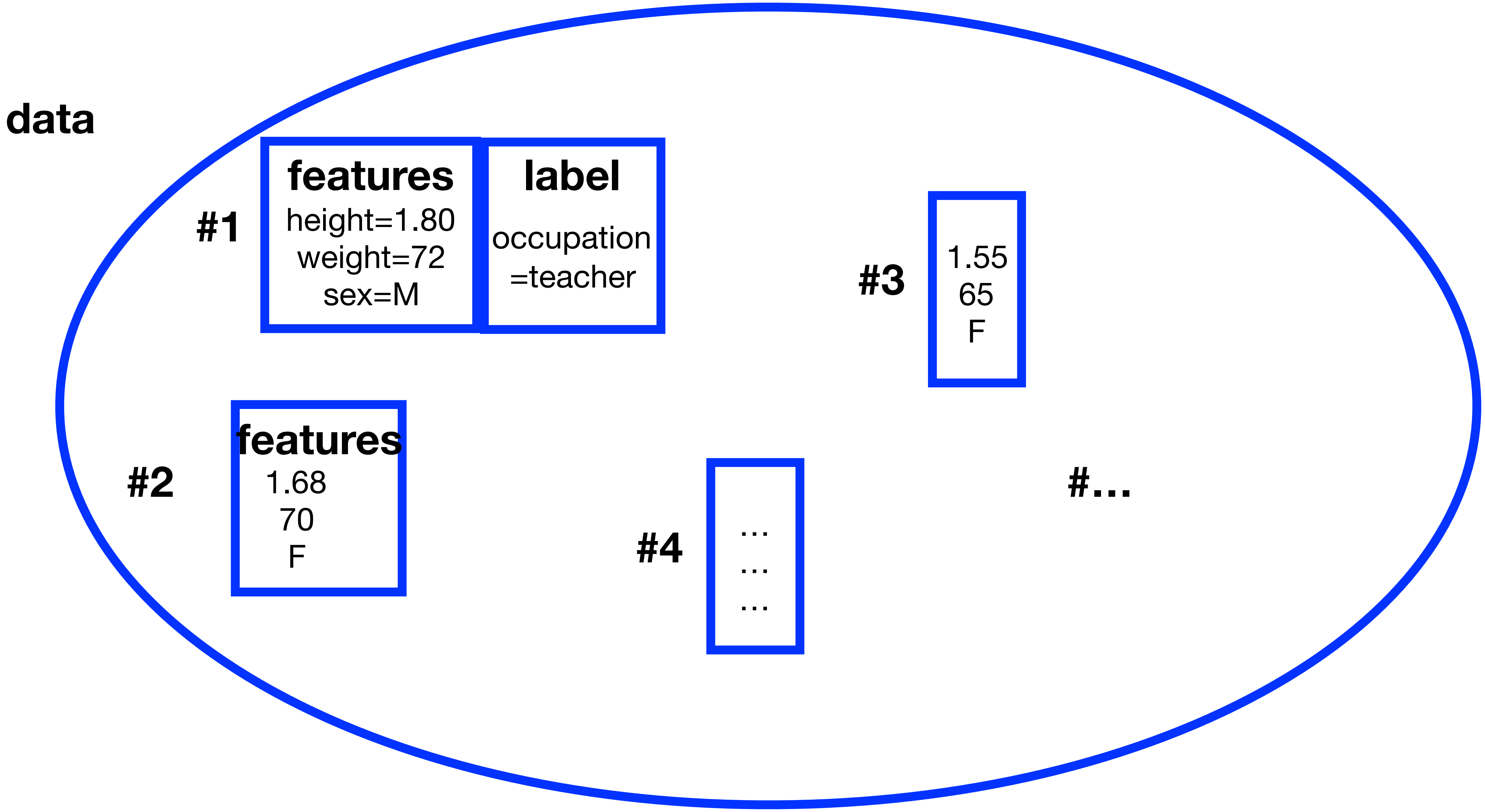
#4
…
…
…
..

#…

Supervised vs. Semi-supervised vs. Unsupervised

Suppose you want to predict occupation based on physical features

but only a few data points have labels



data

#1

features

height=1.80

weight=72

sex=M

label

occupation

=teacher

#2

features

1.68

70

F

#3

1.55

65

F

#4

…

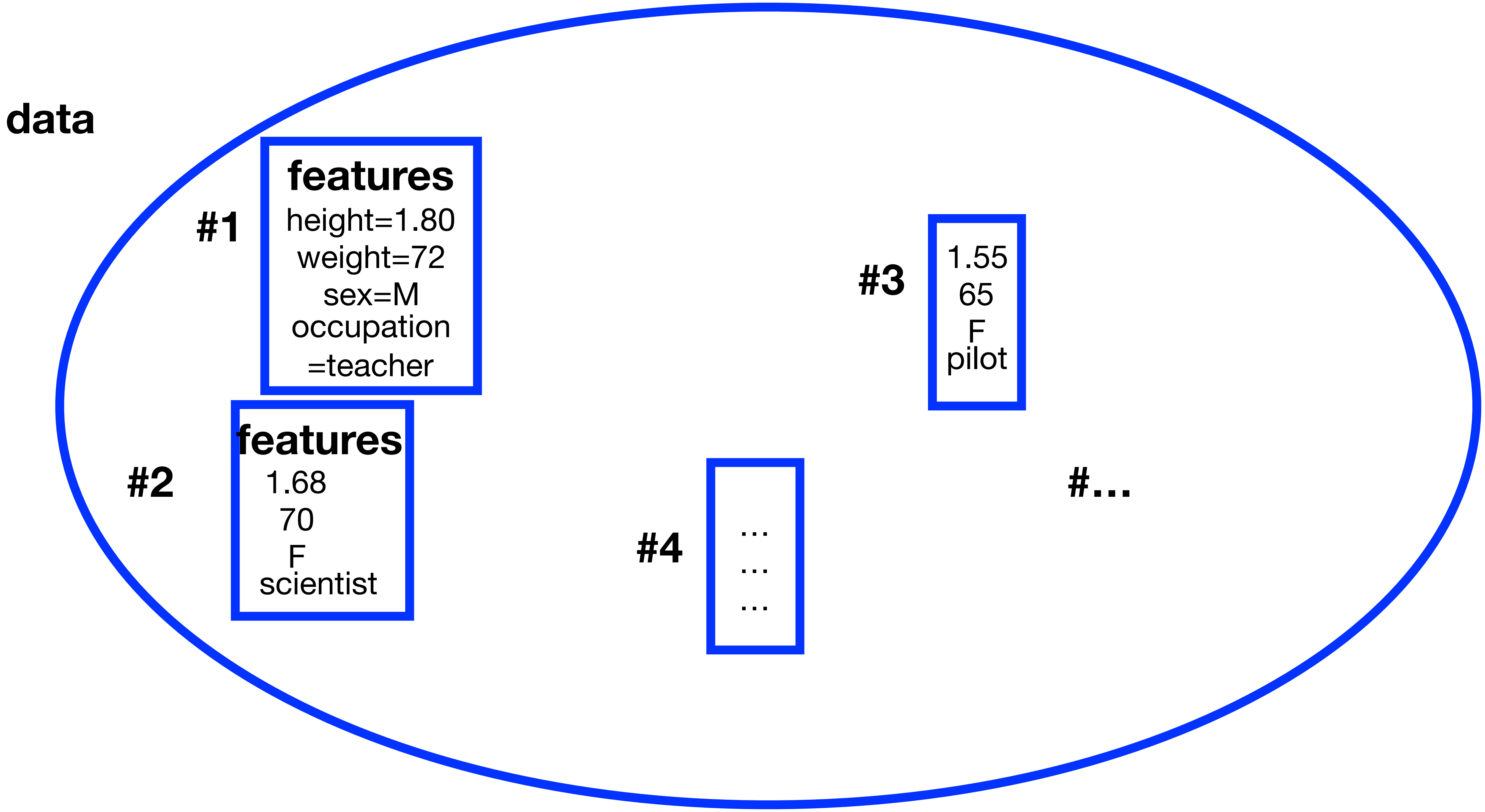
…

…

#…

Supervised vs. Semi-supervised vs. Unsupervised

Suppose you want to understand relationship between occupation and physical features



data

#1

features
height=1.80
weight=72
sex=M
occupation
=teacher

#2

features
1.68
70
F
scientist

#3

1.55
65
F
pilot

#4

…
…
…

#…

Supervised learning

training in progress…

Supervised learning

features

training in progress…

Supervised learning

label

features

training in progress…

Supervised learning

training in progress…

Supervised learning

training in progress…

Supervised learning

training in progress…

Supervised learning

training in progress…

Supervised learning

test!

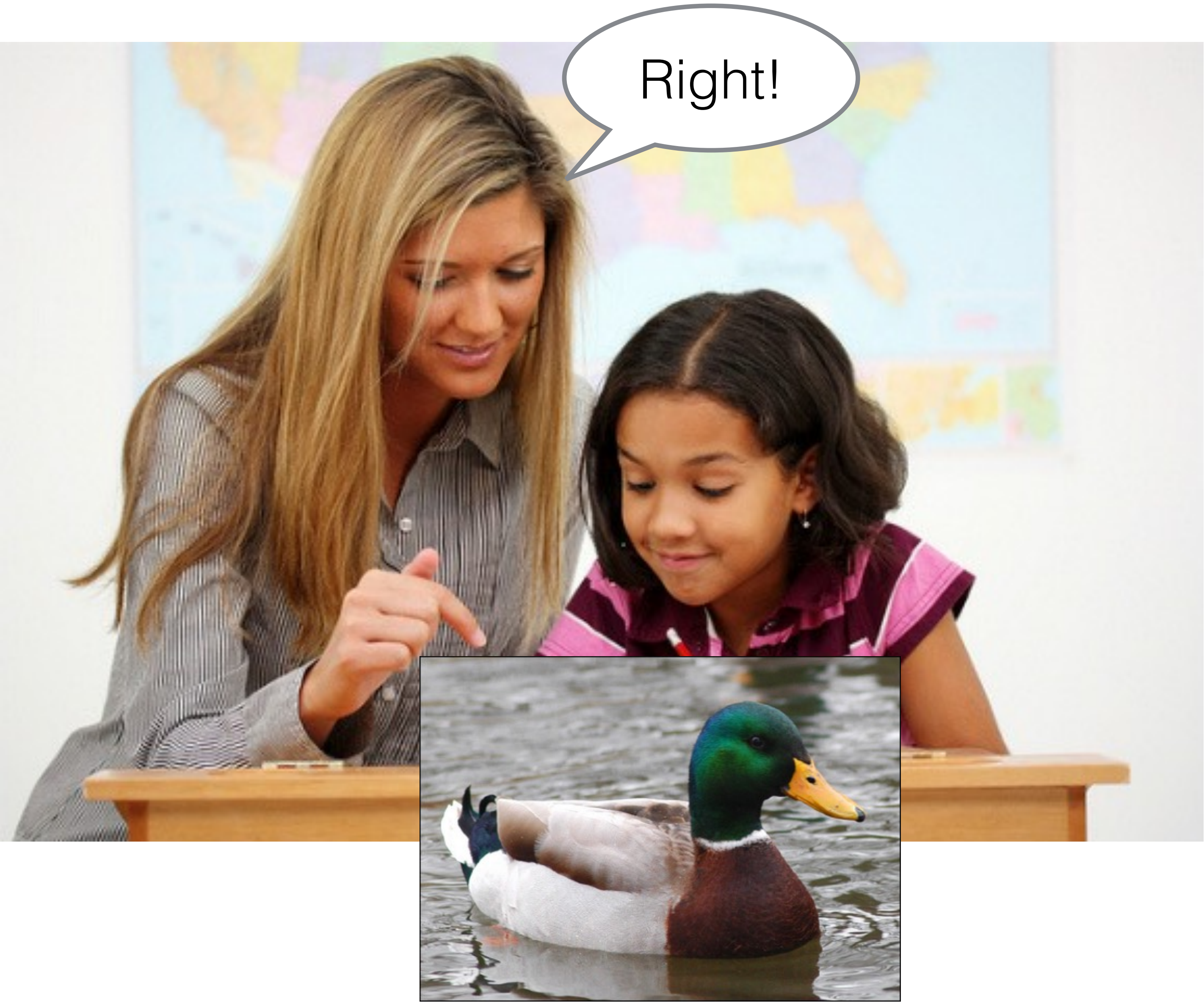
Supervised learning

test!

Supervised learning

success metric

test! accuracy =100%



Unsupervised learning

training in progress…

Unsupervised learning

features

training in progress…

Unsupervised learning

features

training in progress…

Unsupervised learning

training in progress…

Unsupervised learning

training in progress…

Unsupervised learning

training in progress…

Unsupervised learning

training in progress…

Unsupervised learning

training in progress…

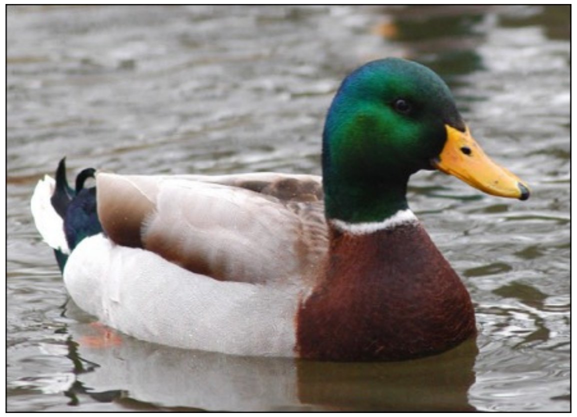
Unsupervised learning

Outcome

Unsupervised learning

Outcome

Ok! I got this. I think there are two types of things:



Unsupervised learning

Outcome

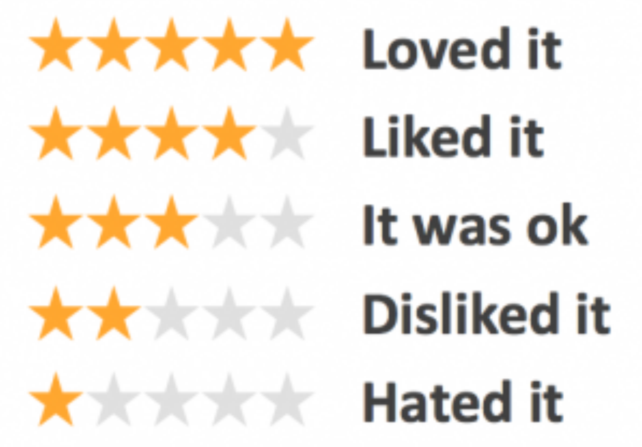
success metric

????

Some vocabulary

Types of features/label:

- continuous: e.g. height, temperature, …
- categorical: e.g. F/M, 0/1, teacher/journalist/doctor (these are called classes)
- ordinal: categorical but with an order between classes (e.g. star ratings)



★★★★★	Loved it
★★★★☆	Liked it
★★★☆☆	It was ok
★★☆☆☆	Disliked it
★☆☆☆☆	Hated it

Supervised tasks:

- label is continuous -> regression
- label is categorical/ordinal -> classification
- more than one label: multi-task

Training

remember:

training

(best)

parameters

data

+

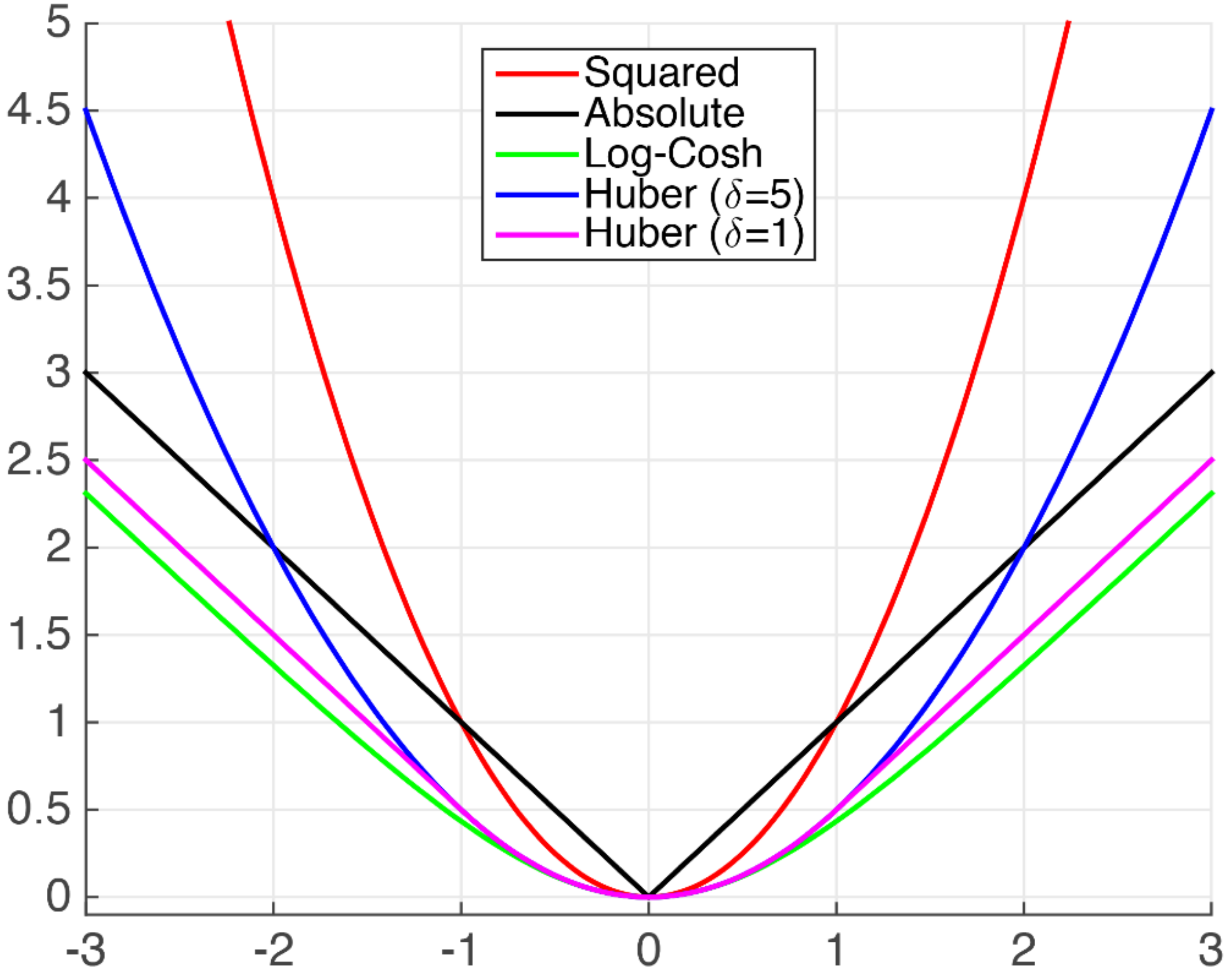
algorithm

How do we define “best”??

Loss/cost function

introduce a function Loss of the parameters

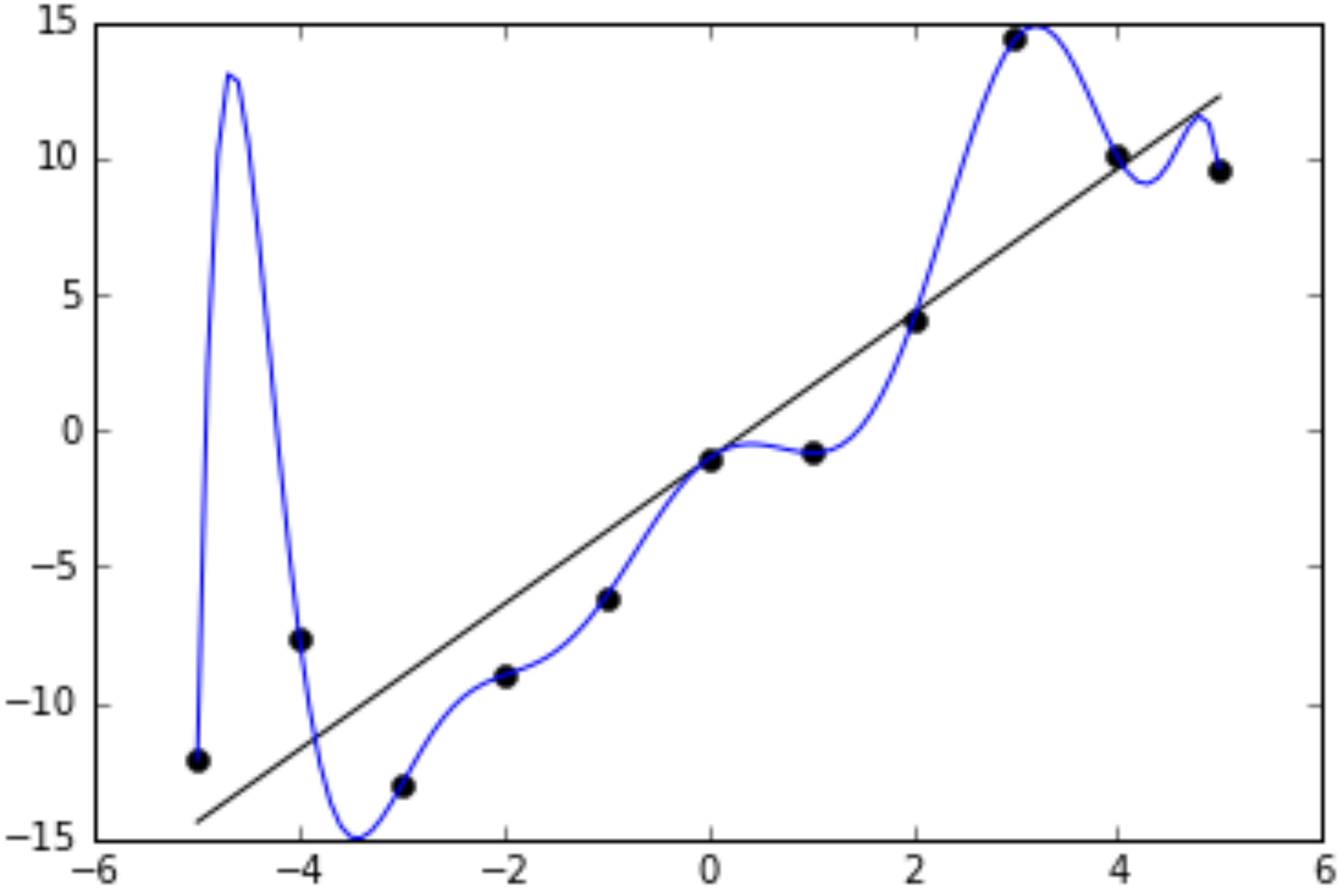
Loss(bad parameters) = very high



Loss(best parameters) = lowest

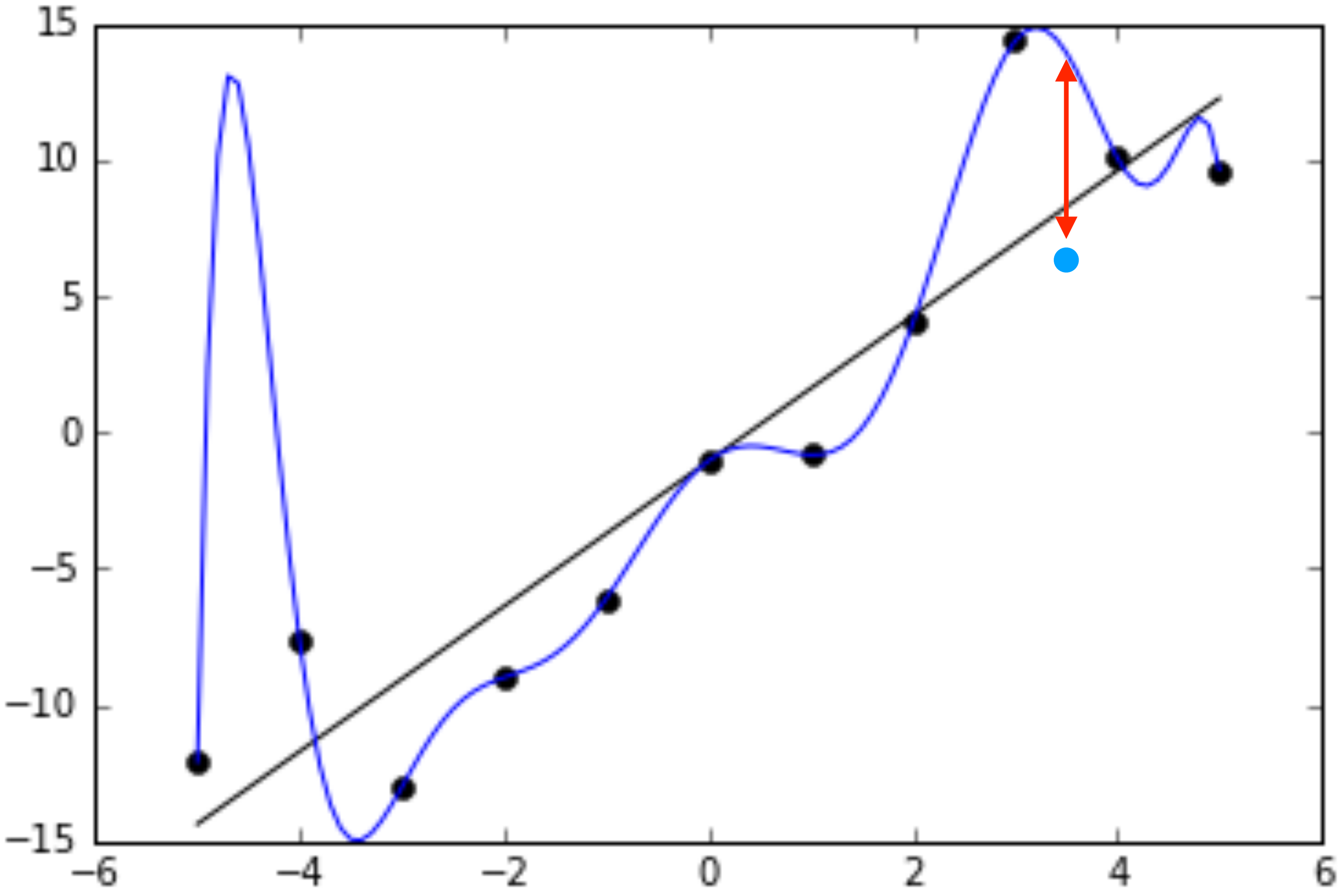
Overfitting

Problem: what will happen if I try my model on new data never used in training?



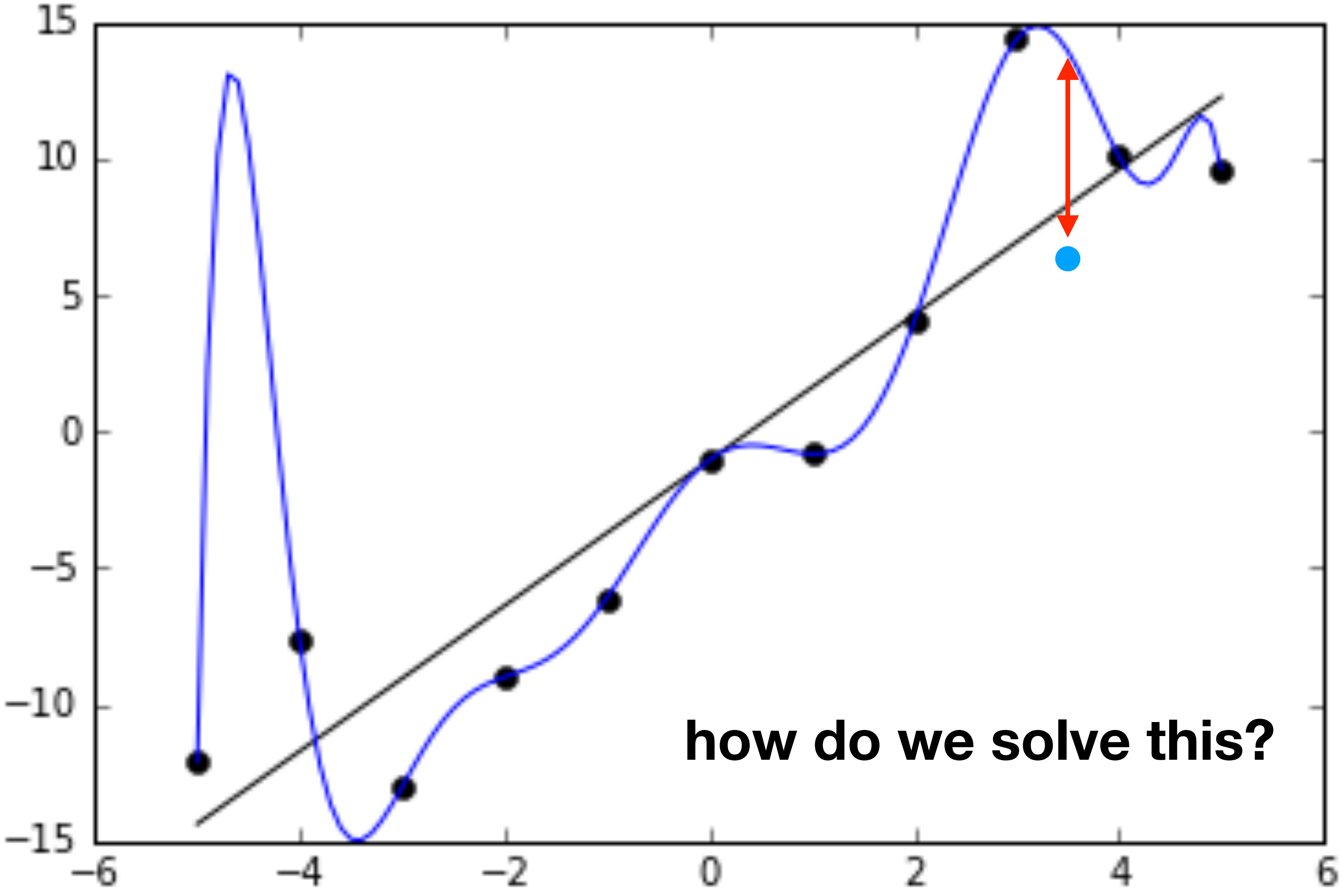
Overfitting

Problem: what will happen if I try my model on new data never used in training?



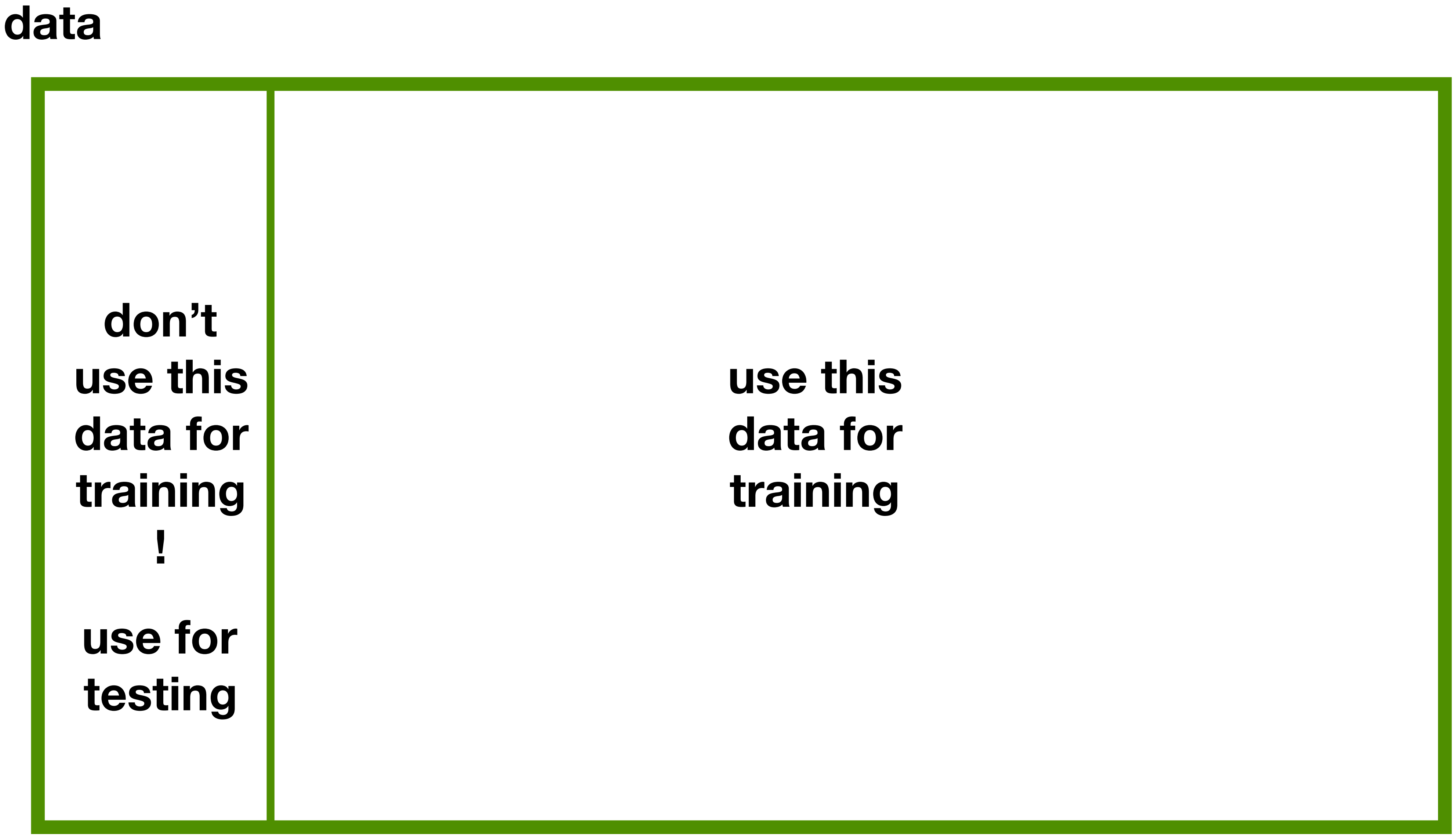
Overfitting

Problem: what will happen if I try my model on new data never used in training?



how do we solve this?

Validation



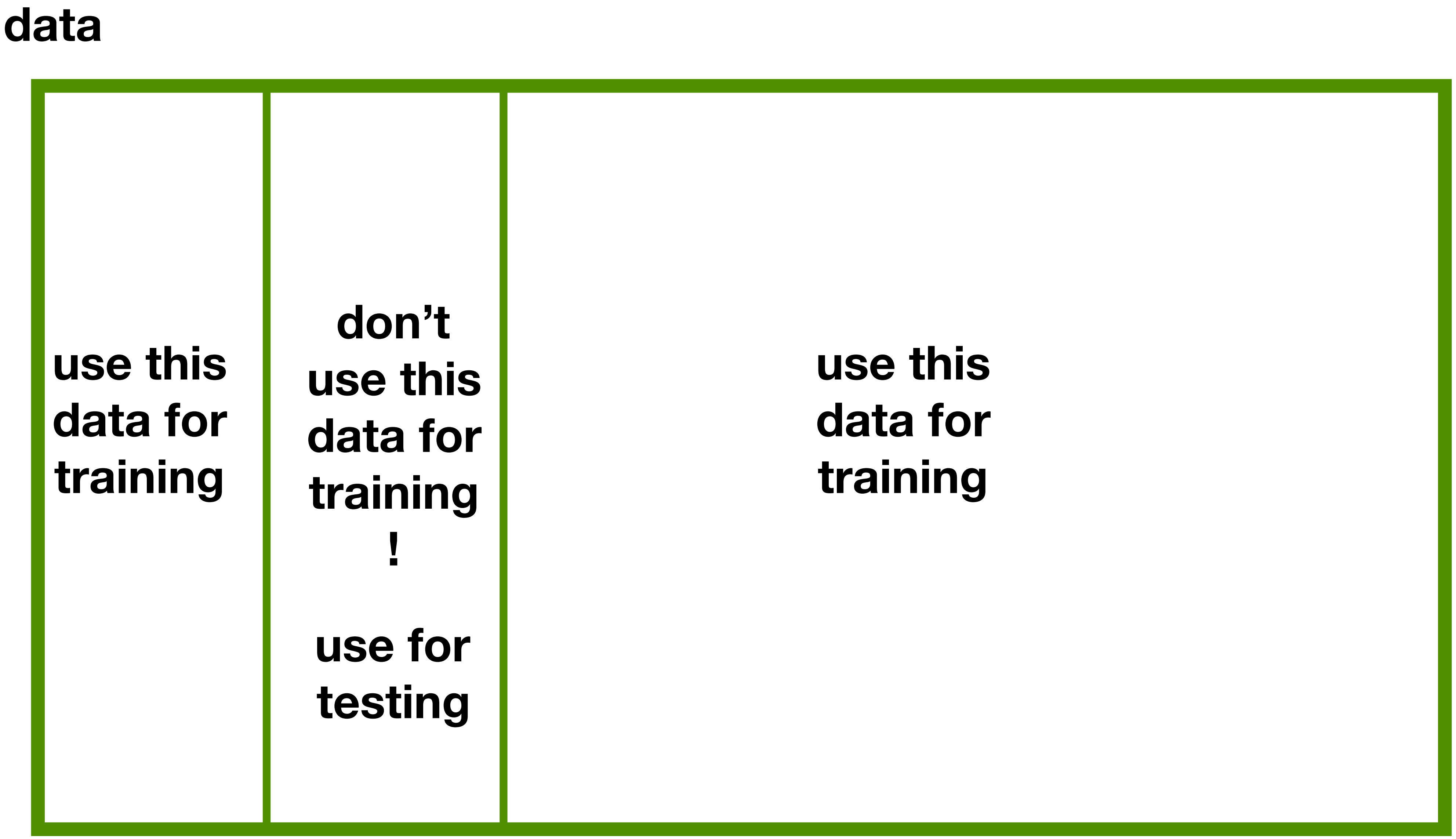
data

don't use this data for training !

use for testing

use this data for training

Validation



data

use this data for training

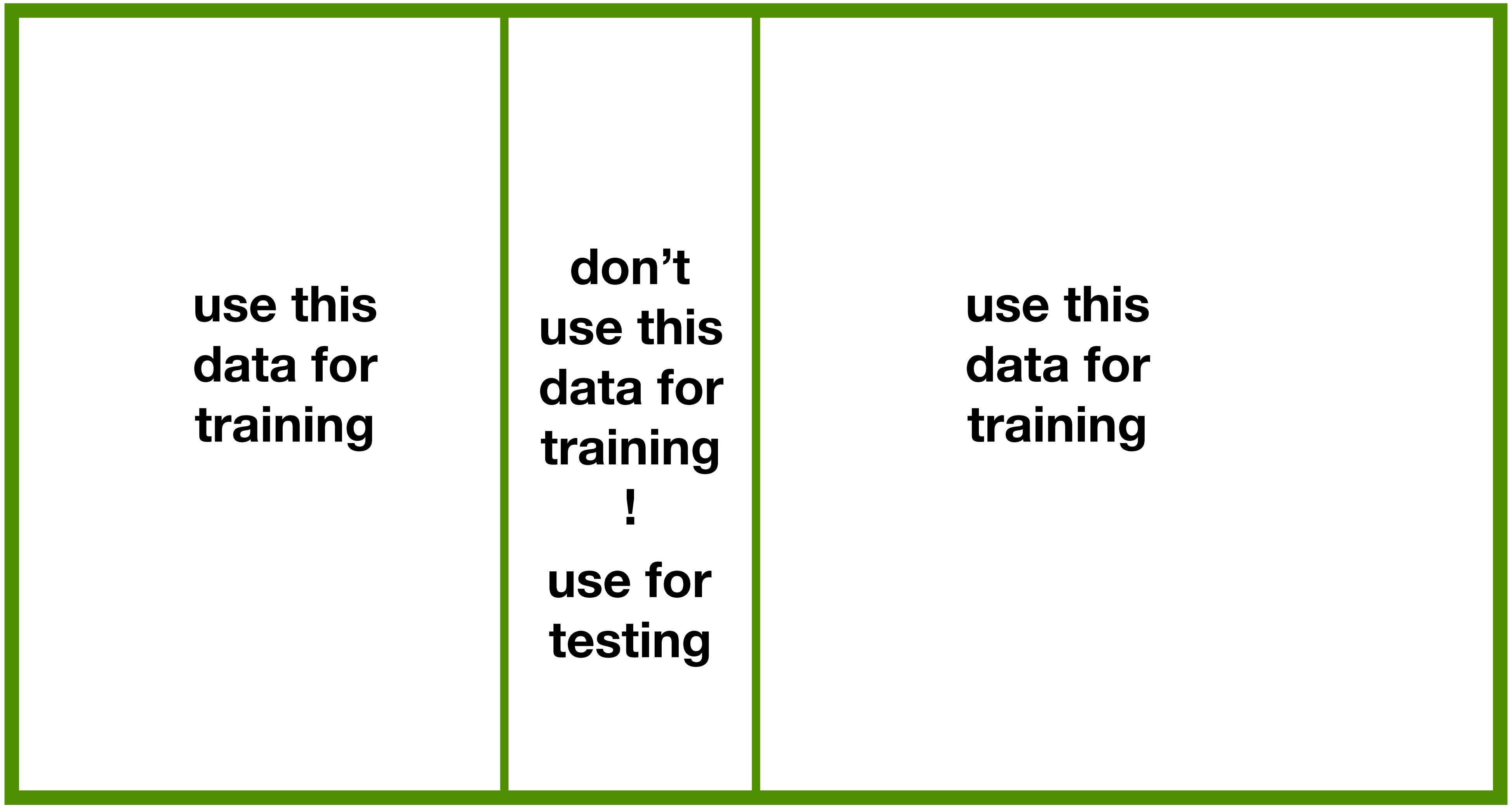
don't use this data for training !

use for testing

use this data for training

Validation

data



use this data for training

don't use this data for training !

use for testing

use this data for training

…k-fold validation

Validation

take an average of the testing performance over the k times

↓

choose the hyper parameters that make this average best!