Statistics and Machine Learning

Model complexity and cross validation

Week 6 02/22 — 02/26 2021 Spring

Contents of Week 6

Main concept: Bias-variance tradeoff

- Review of our 5 basic blocks in supervised learning
- Model complexity
- THE most important parameter of ML model: test error
- Ways to obtain test error
- Bias, variance, and the tradeoff

Review:

5 important elements in supervised machine learning

Data set

model

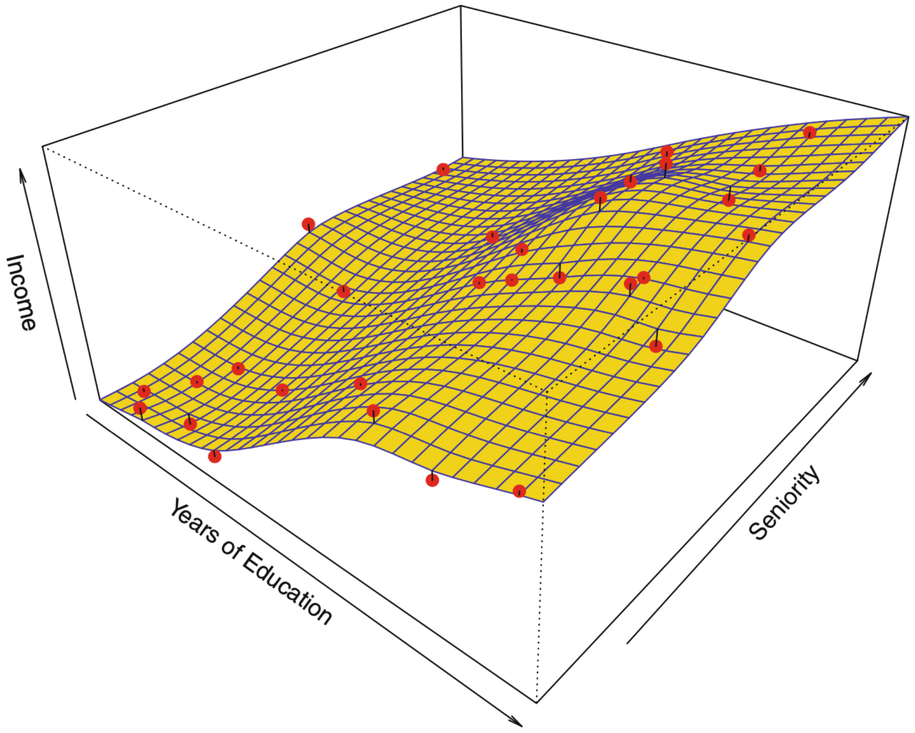
Loss function

Training error

Test error

Data set

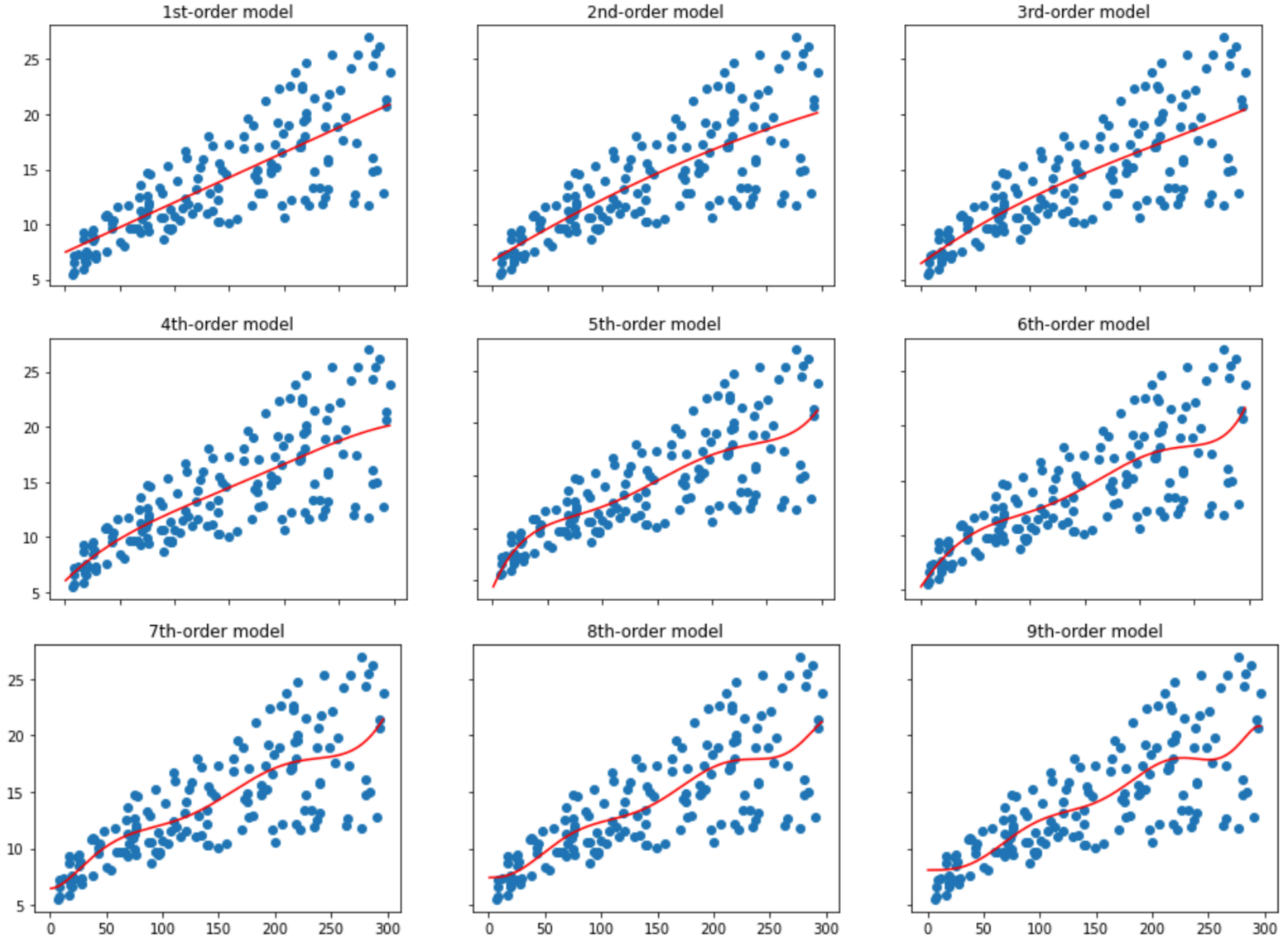
Data set: tabular information among which we want to find out the mutual relation.



https://github.com/JWarmenhoven/ISLR-python/tree/master/Notebooks/Data

Model

Model: the mathematical relation which you believe to exist within data set



How to use quadratic model?

TV
230.1
44.5
17.2
151.5
180.8

X

Sales
22.1
10.4
9.3
18.5
12.9

Y

The linear mode:

$$f(x) = ax + b$$

Model parameters:

$$\{a, b\}$$

How to use quadratic model?

TV
230.1
44.5
17.2
151.5
180.8

X

Sales
22.1
10.4
9.3
18.5
12.9

Y

The linear mode:

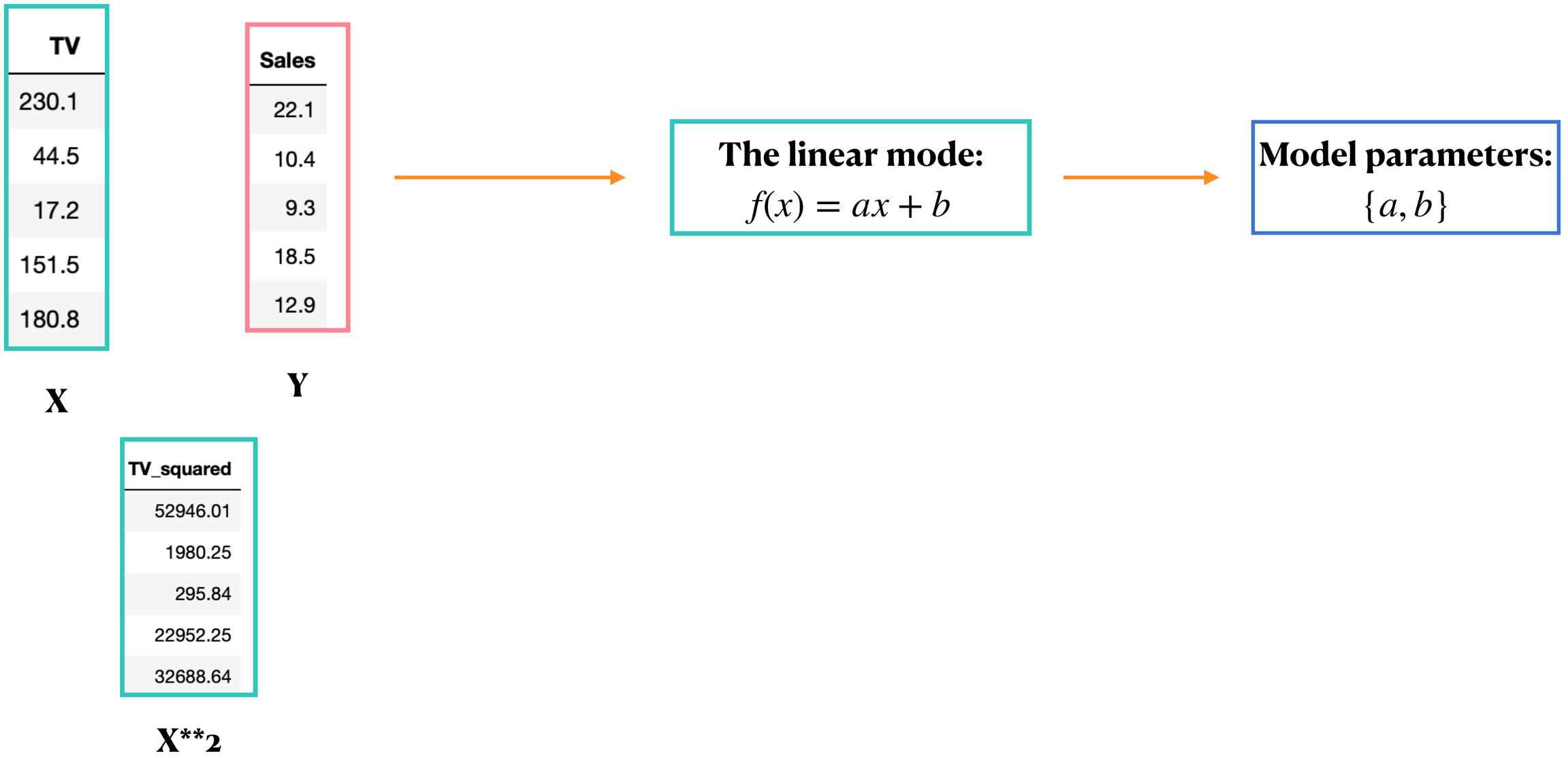
$$f(x) = ax + b$$

Model parameters:

$$\{a, b\}$$

TV_squared
52946.01
1980.25
295.84
22952.25
32688.64

How to use quadratic model?



How to use quadratic model?

TV
230.1
44.5
17.2
151.5
180.8

X

Sales
22.1
10.4
9.3
18.5
12.9

Y

TV_squared
52946.01
1980.25
295.84
22952.25
32688.64

X**2

The linear mode:

$$f(x) = ax + b$$

Model parameters:

$$\{a, b\}$$

The quadratic mode:

$$f(x) = ax^2 + bx + c$$

How to use quadratic model?

TV
230.1
44.5
17.2
151.5
180.8

X

Sales
22.1
10.4
9.3
18.5
12.9

Y

TV_squared
52946.01
1980.25
295.84
22952.25
32688.64

X**2

The linear mode:

$$f(x) = ax + b$$

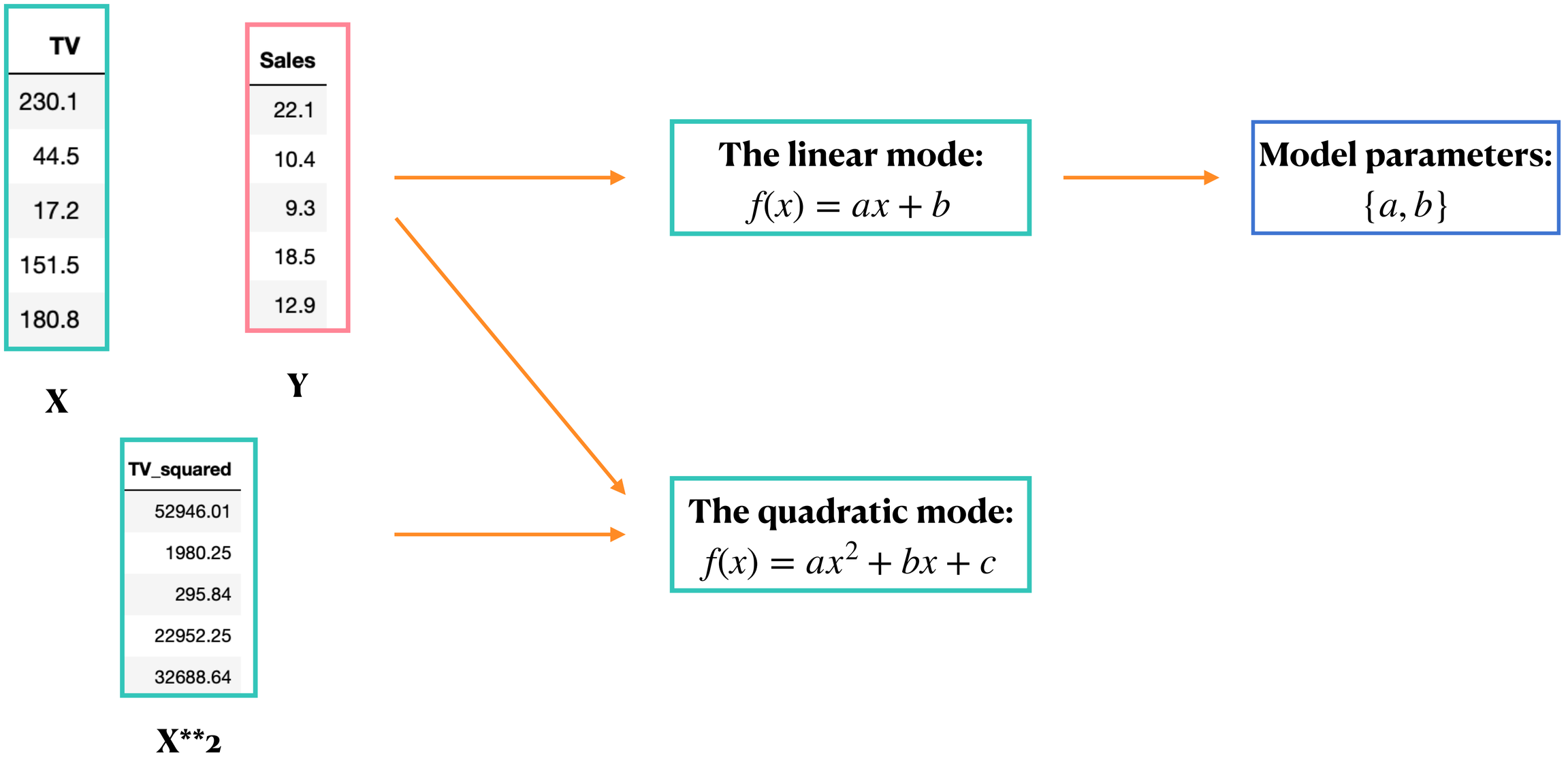
Model parameters:

$$\{a, b\}$$

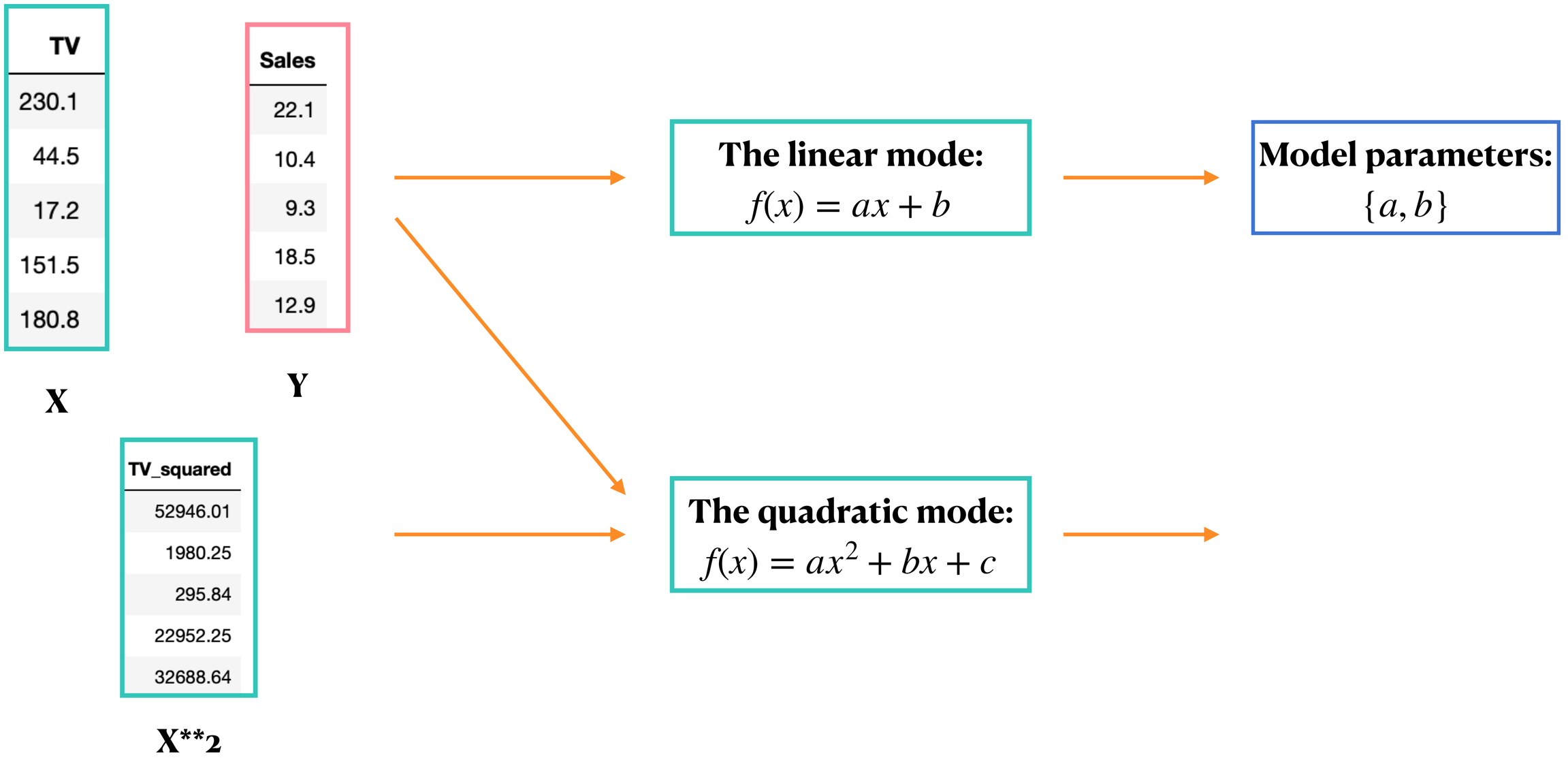
The quadratic mode:

$$f(x) = ax^2 + bx + c$$

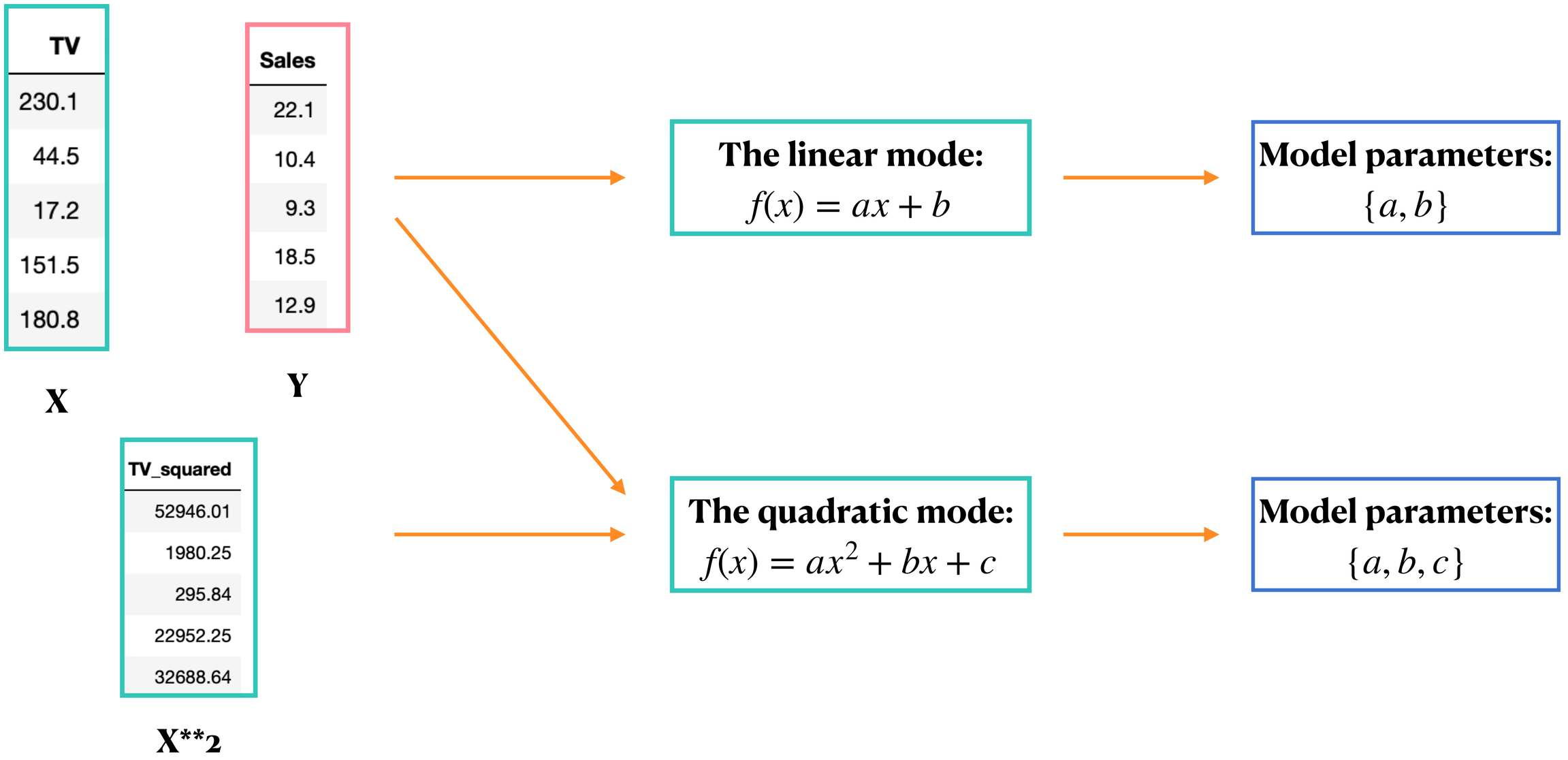
How to use quadratic model?



How to use quadratic model?



How to use quadratic model?



Loss function

Loss function: the error representing how your model perform with data

In regression tasks, we use the mean squared error:

$$MSE := \frac{1}{N}\sum_{i=1}^{N}[f(x_i) - y_y]^2$$

There is also the mean absolute loss: MAE

$$MAE := \frac{1}{N}\sum_{i=1}^{N}|f(x_i) - y_i|$$

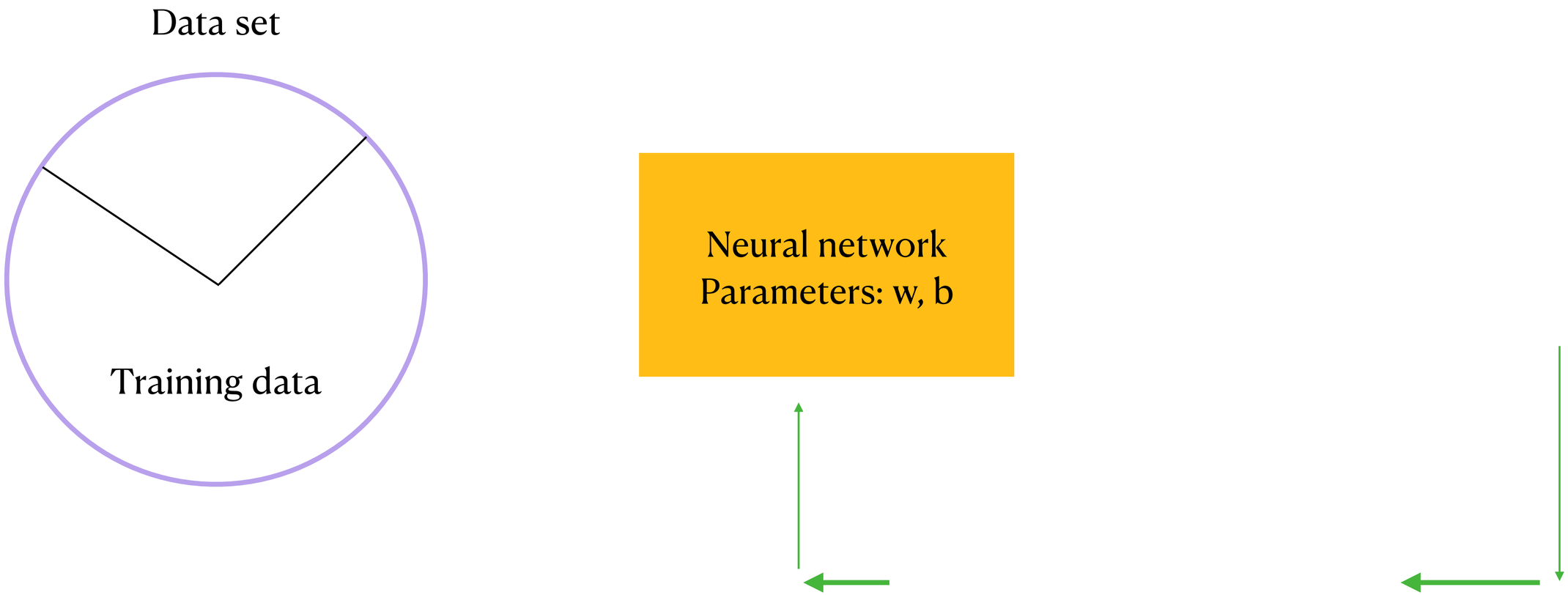
And the cross-entropy loss: when it comes to binary data (y= 0 or 1)

$$CE := \frac{1}{N}\sum_{i=1}^{N}[y_i \log f(x_i) + (1 - y_i)\log(1 - f(x_i))]$$

https://heartbeat.fritz.ai/5-regression-loss-functions-all-machine-learners-should-know-4fb140e9d4b0

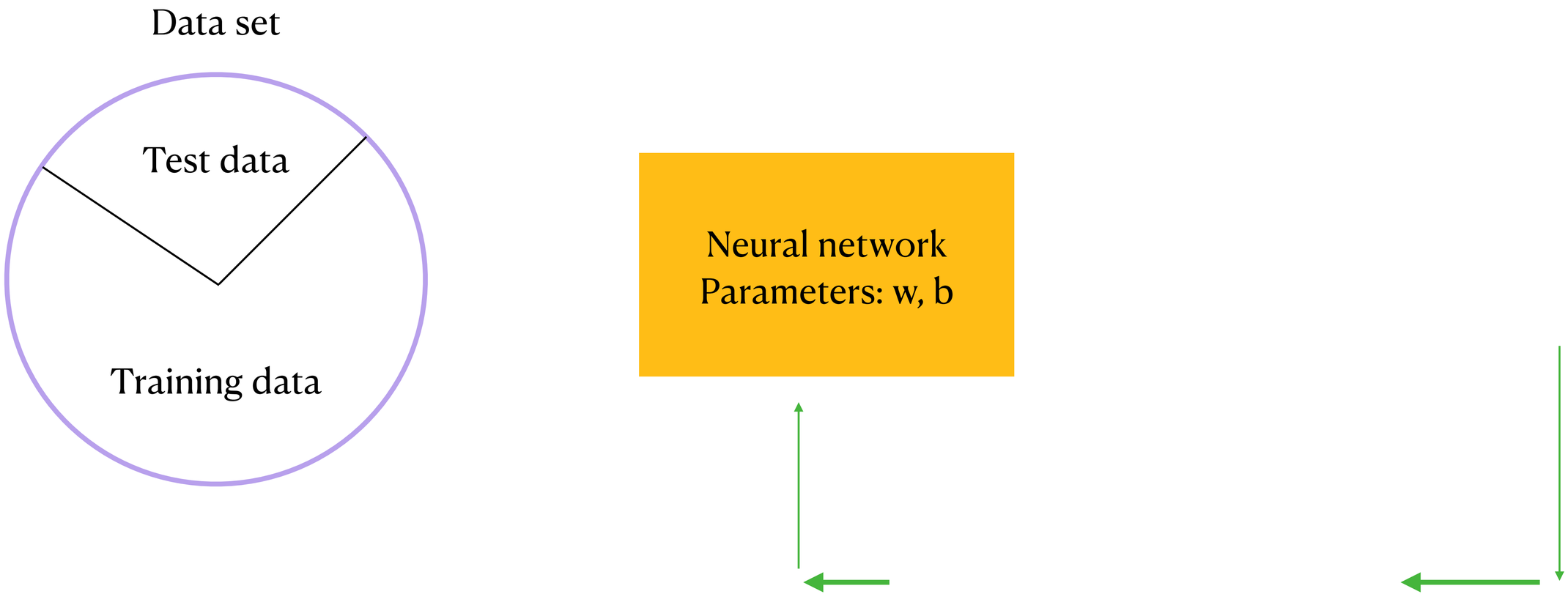
So what is machine learning?

What we want? a capable model, and an unbiased estimate of error rate.



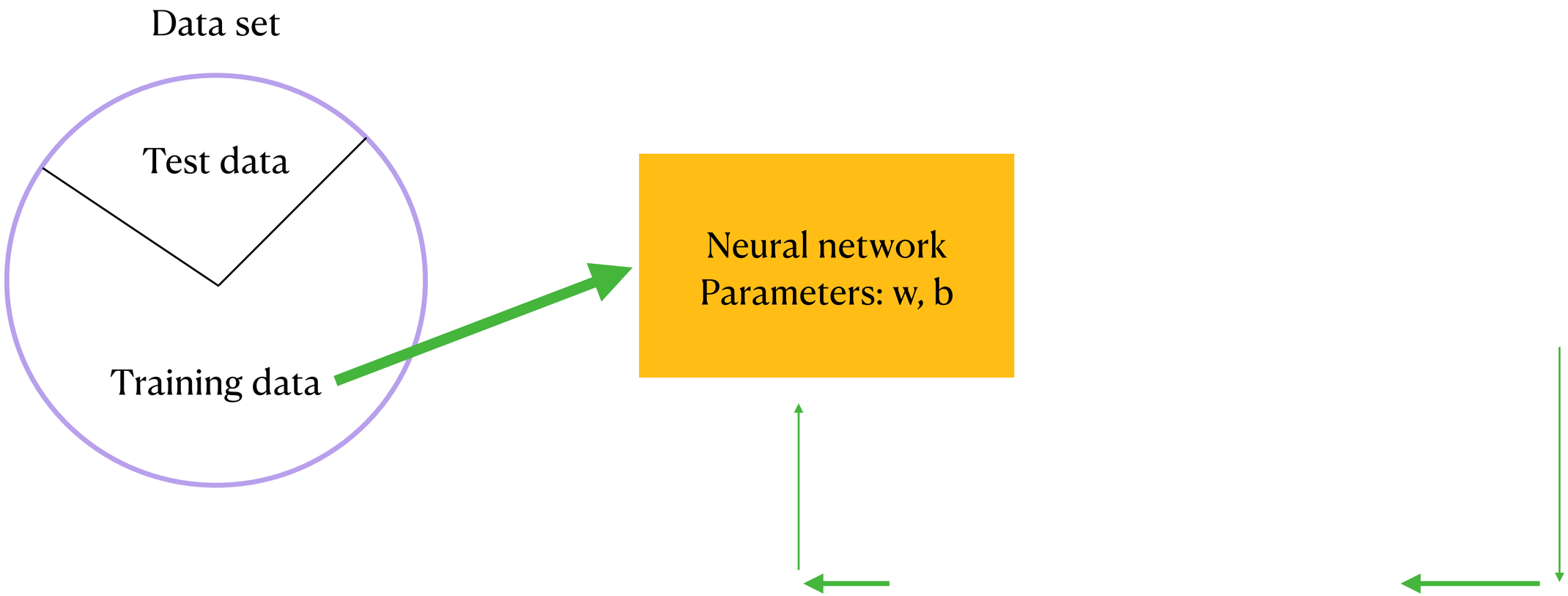
So what is machine learning?

What we want? a capable model, and an unbiased estimate of error rate.



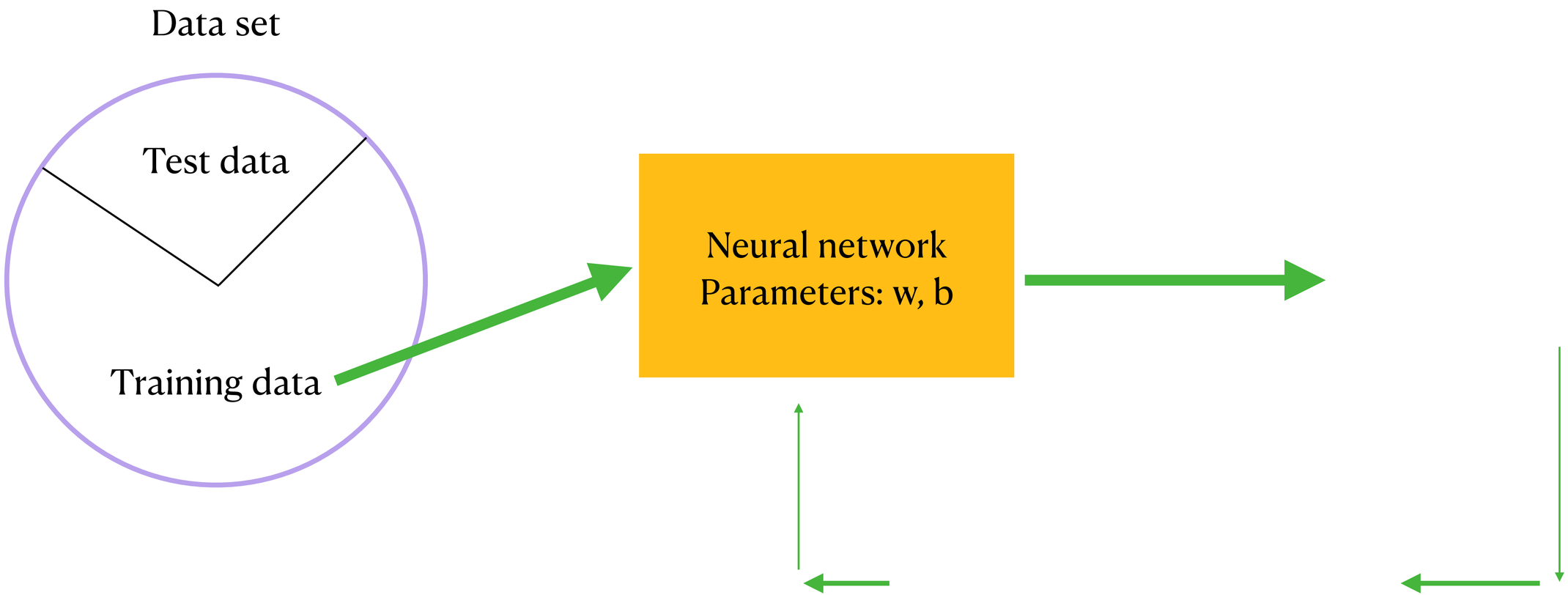
So what is machine learning?

What we want? a capable model, and an unbiased estimate of error rate.



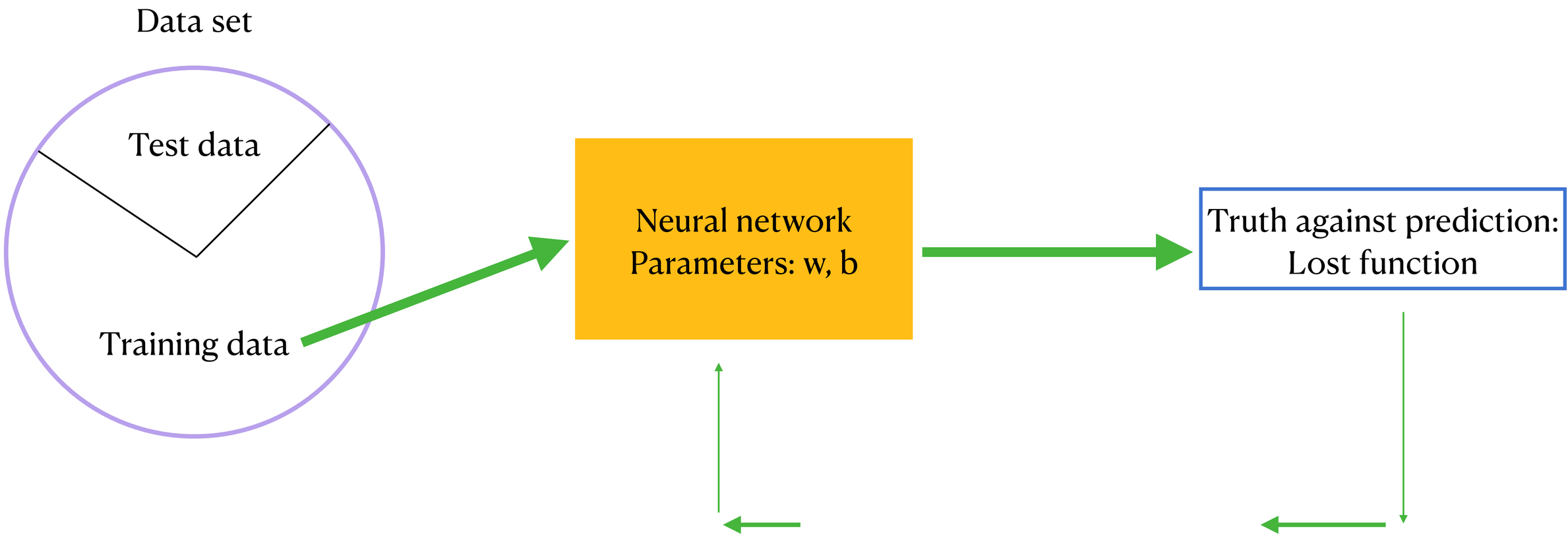
So what is machine learning?

What we want? a capable model, and an unbiased estimate of error rate.



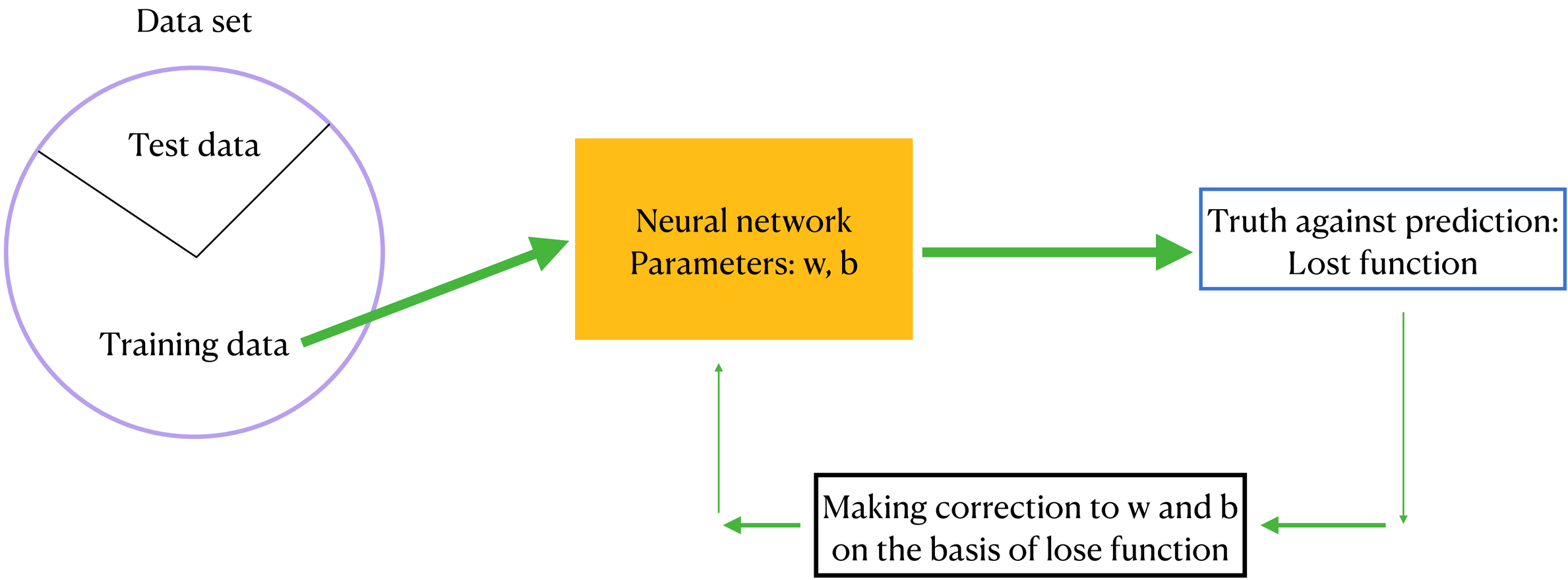
So what is machine learning?

What we want? a capable model, and an unbiased estimate of error rate.



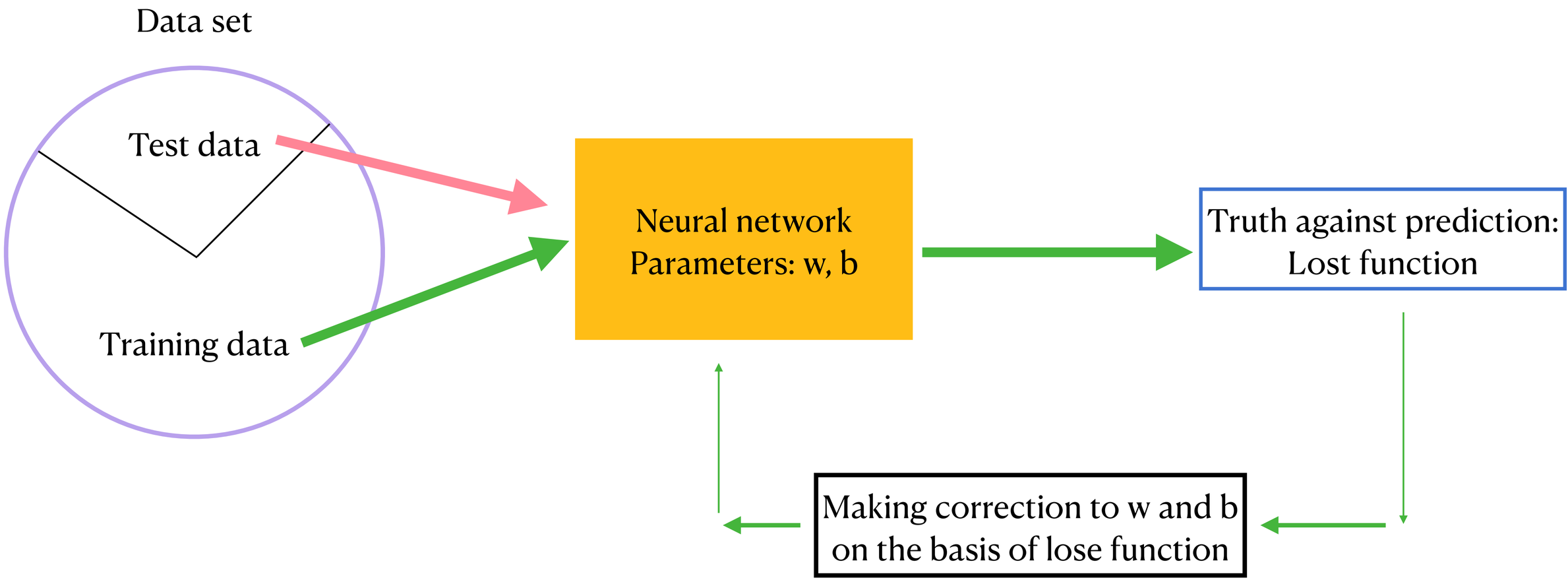
So what is machine learning?

What we want? a capable model, and an unbiased estimate of error rate.



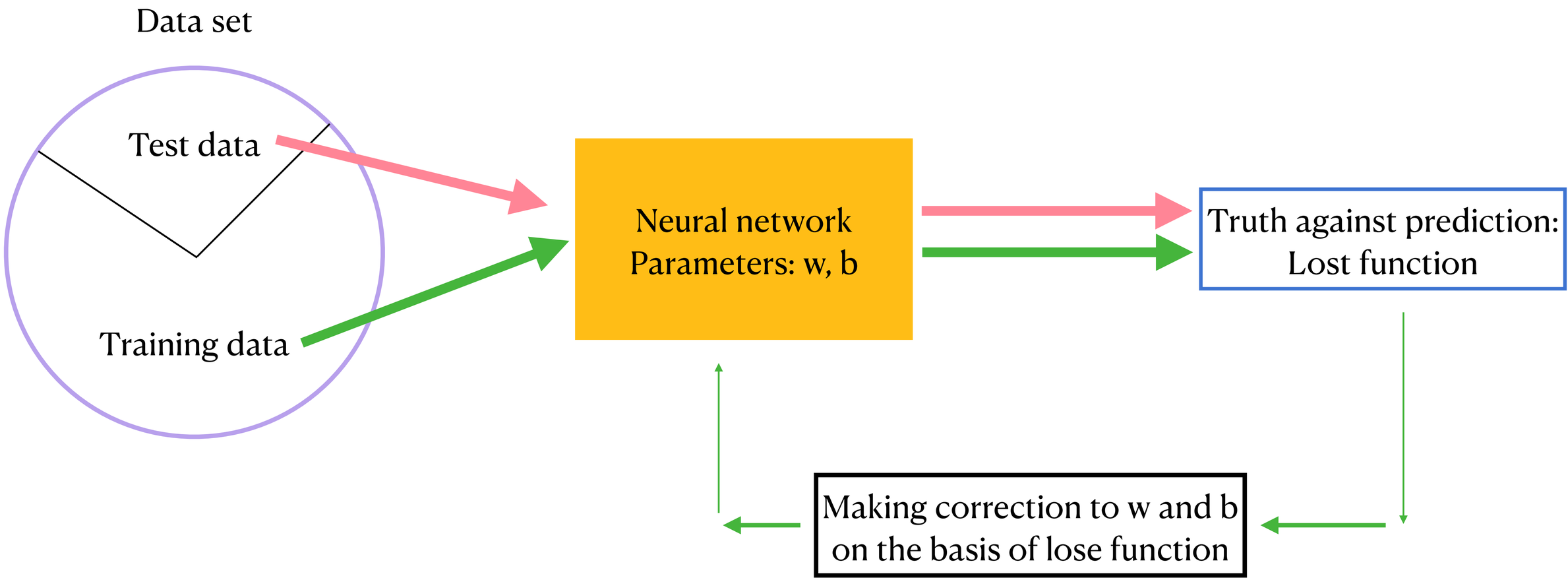
So what is machine learning?

What we want? a capable model, and an unbiased estimate of error rate.



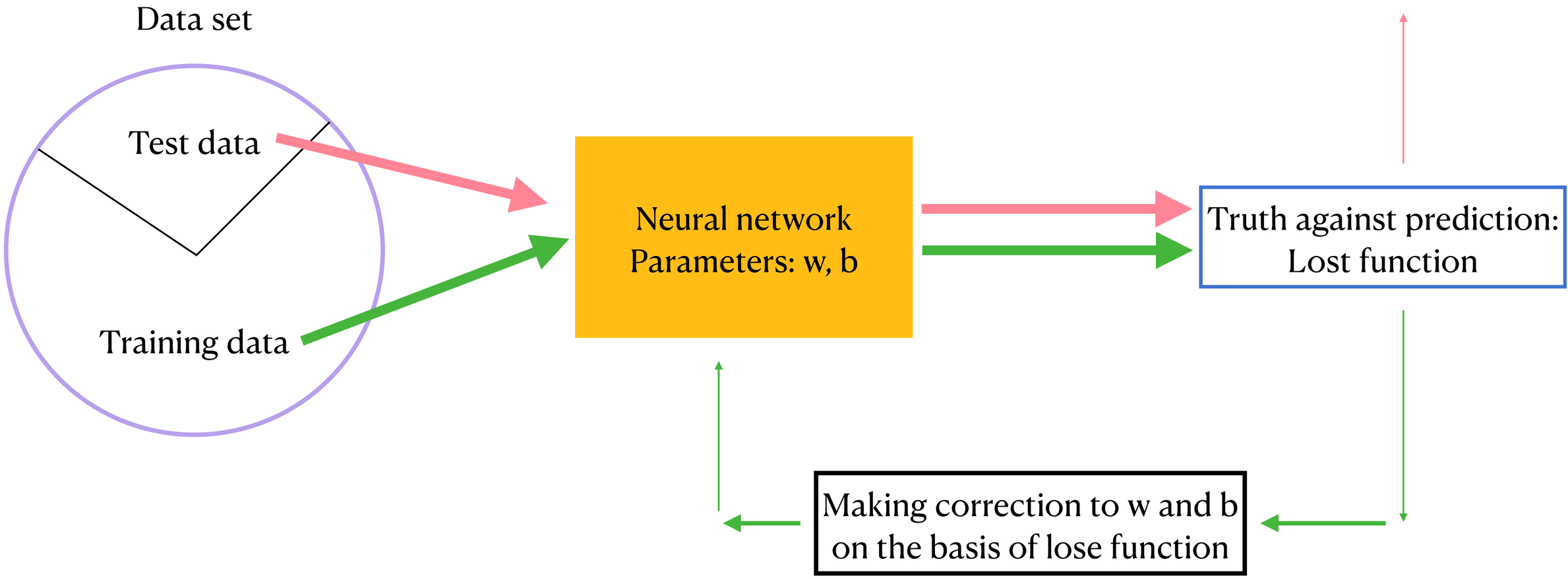
So what is machine learning?

What we want? a capable model, and an unbiased estimate of error rate.



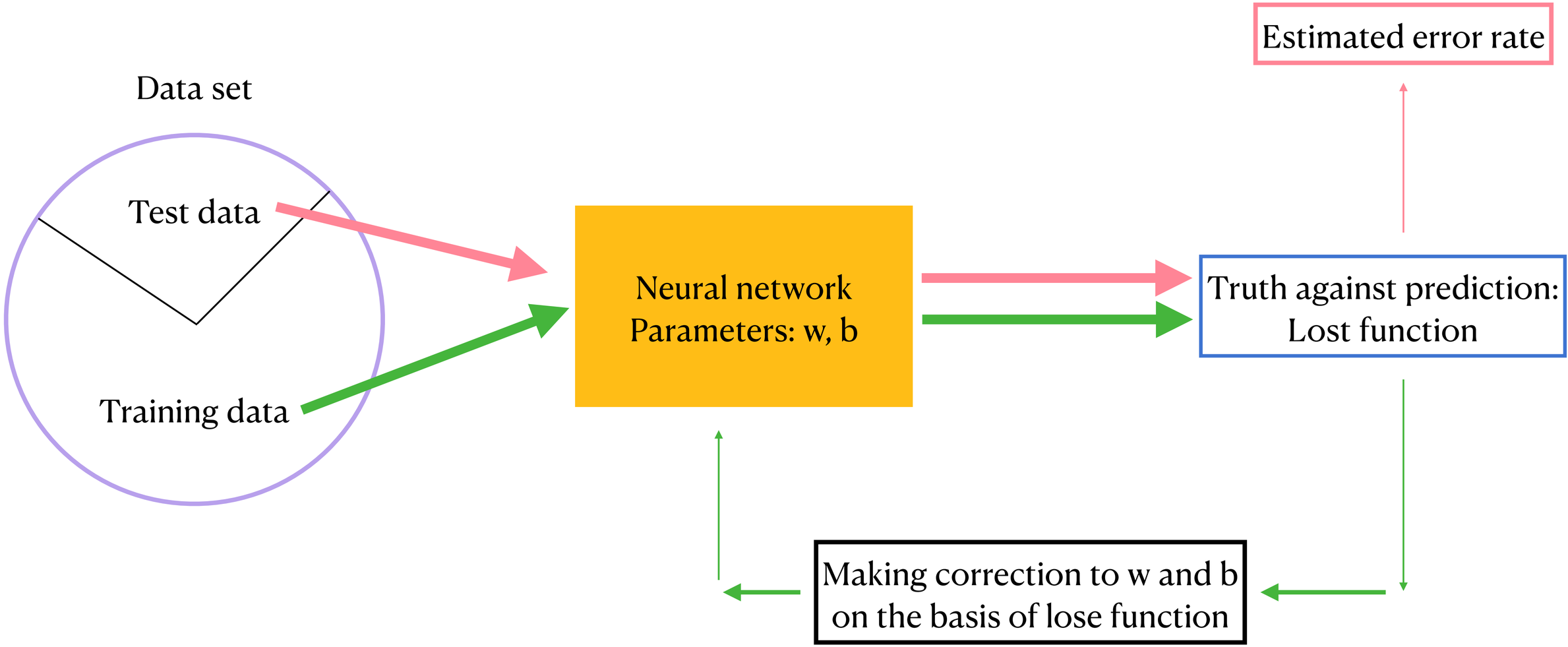
So what is machine learning?

What we want? a capable model, and an unbiased estimate of error rate.



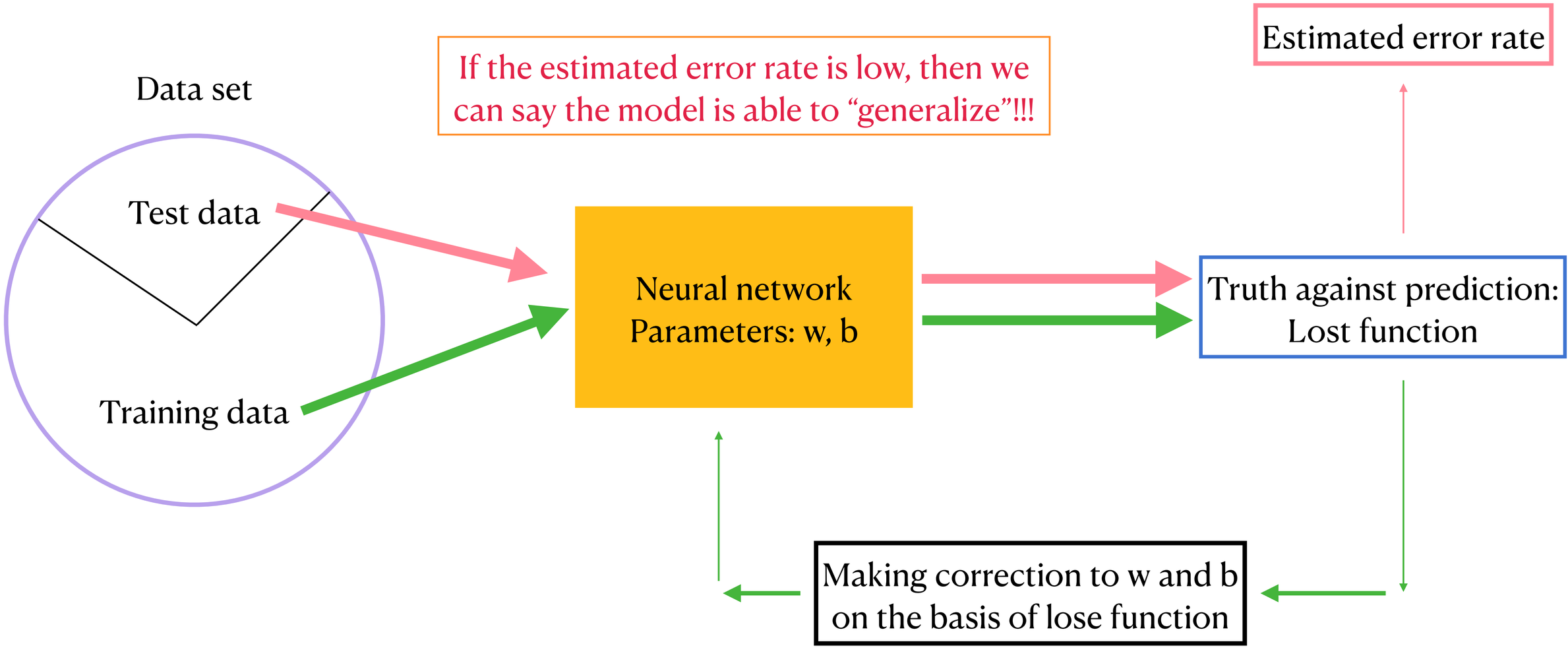
So what is machine learning?

What we want? a capable model, and an unbiased estimate of error rate.

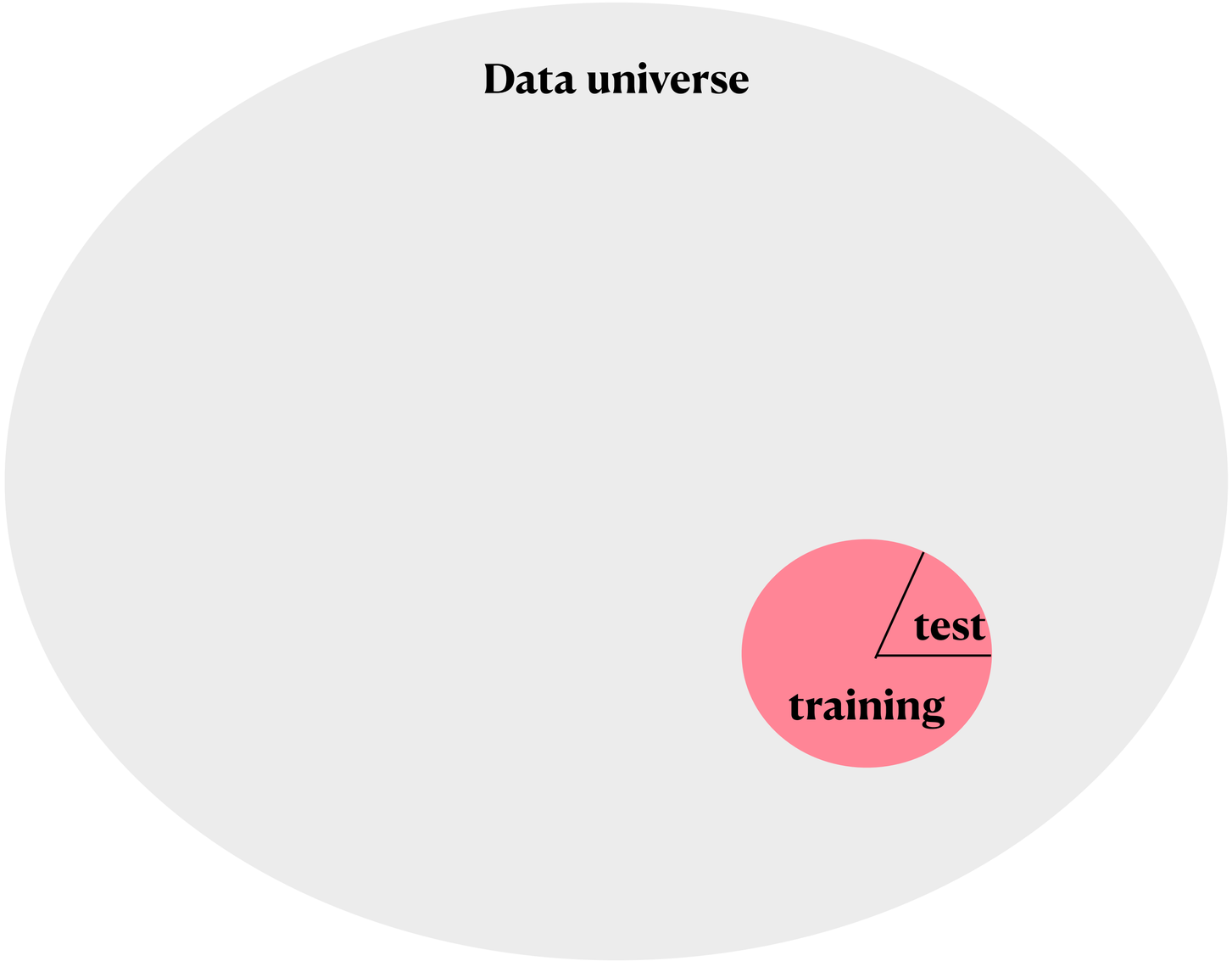


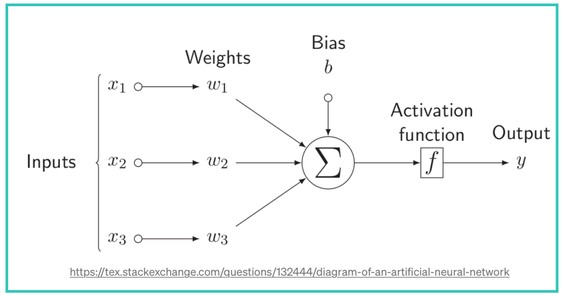
So what is machine learning?

What we want? a capable model, and an unbiased estimate of error rate.



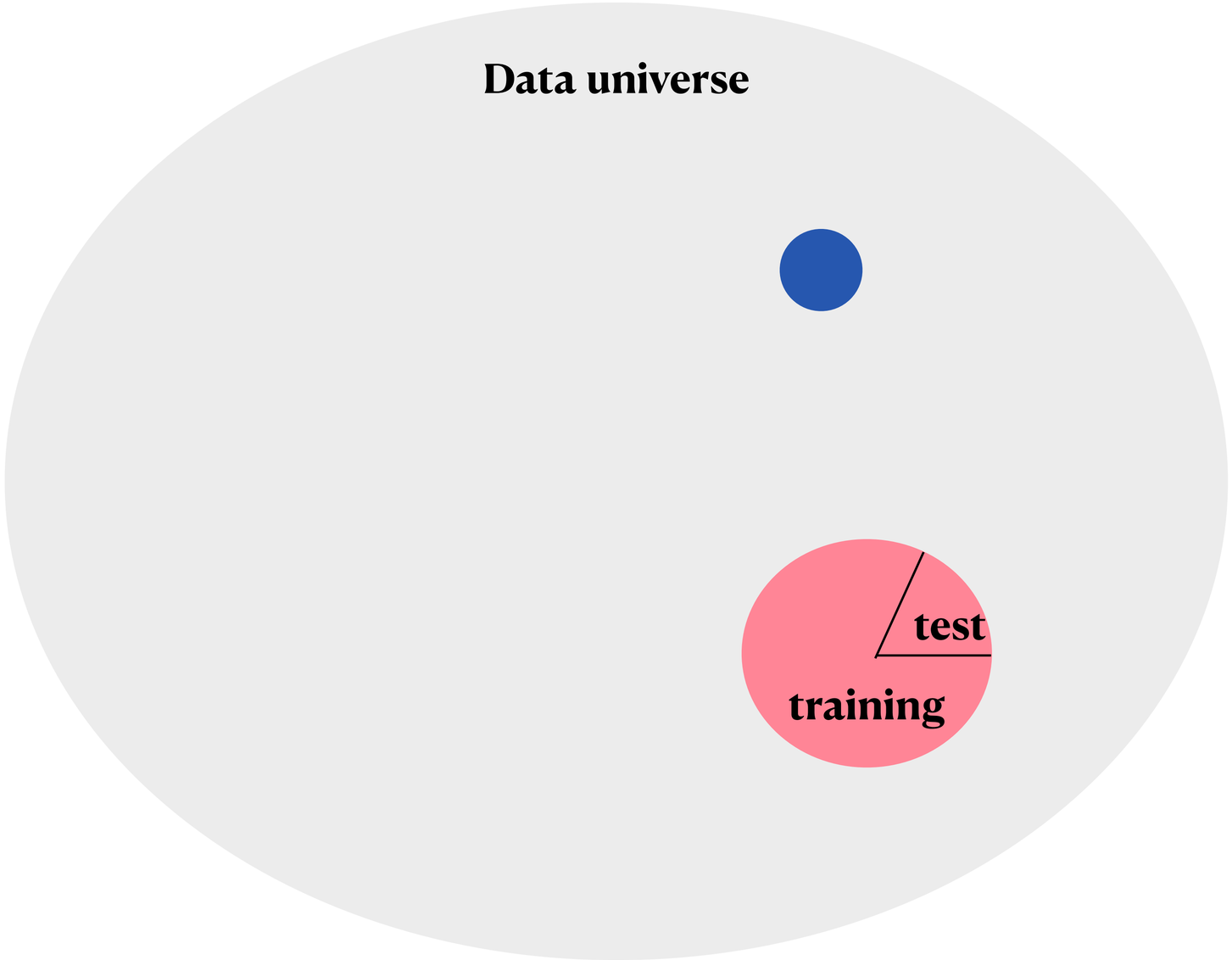
Ultimate goal of ML: deal with unseen data

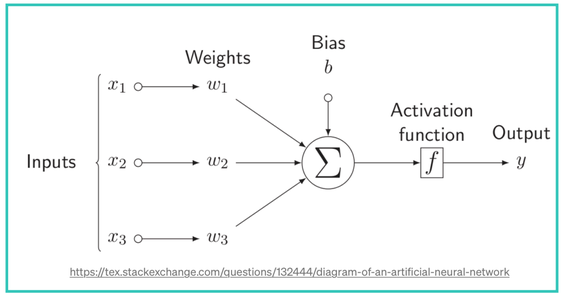




https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network

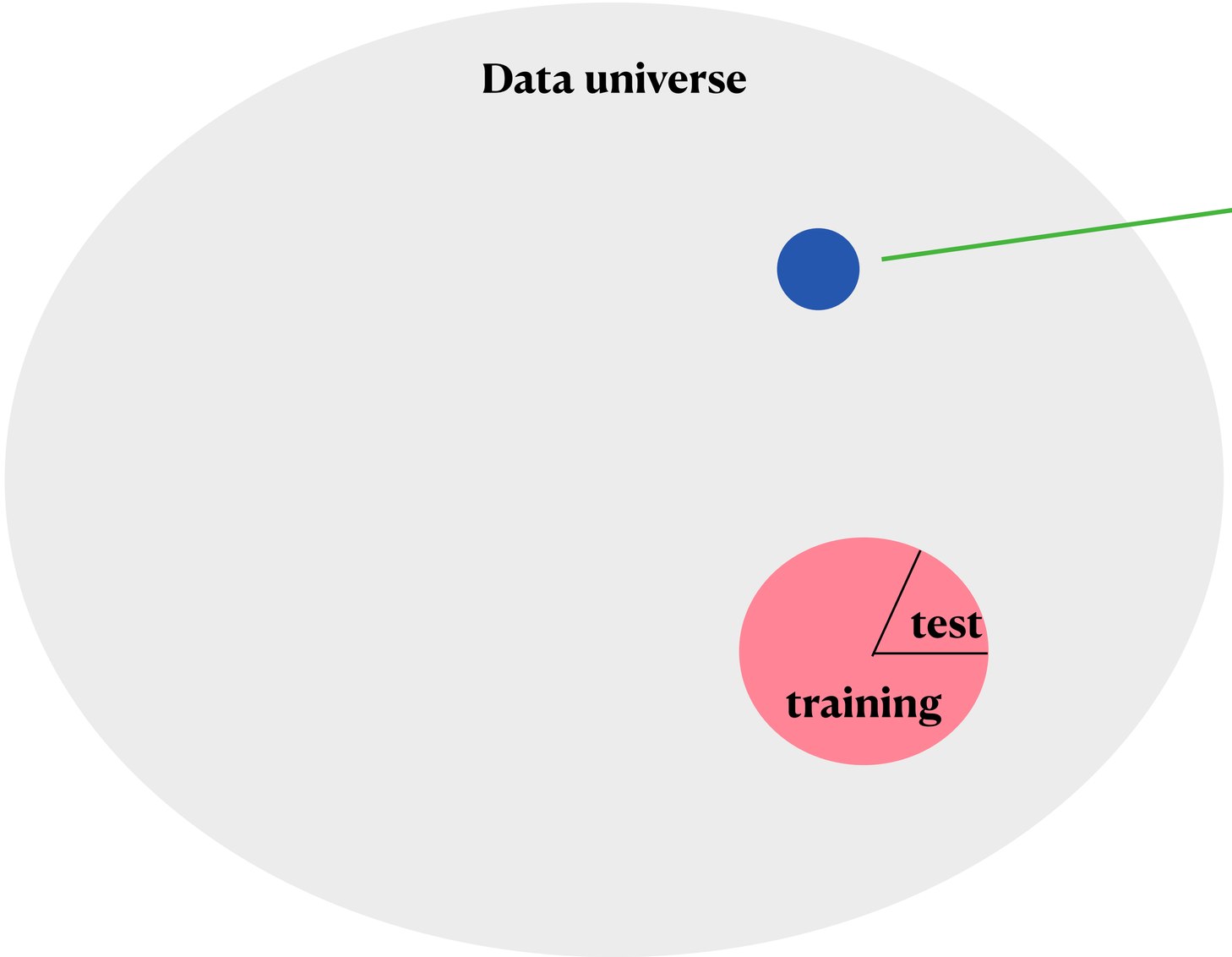
Ultimate goal of ML: deal with unseen data

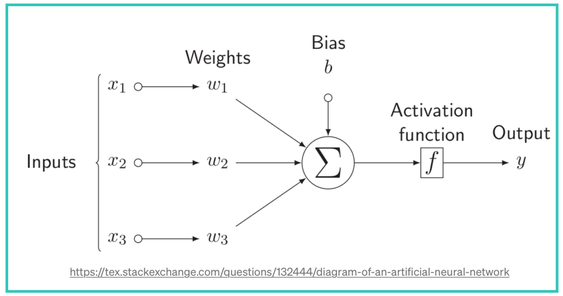




https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network

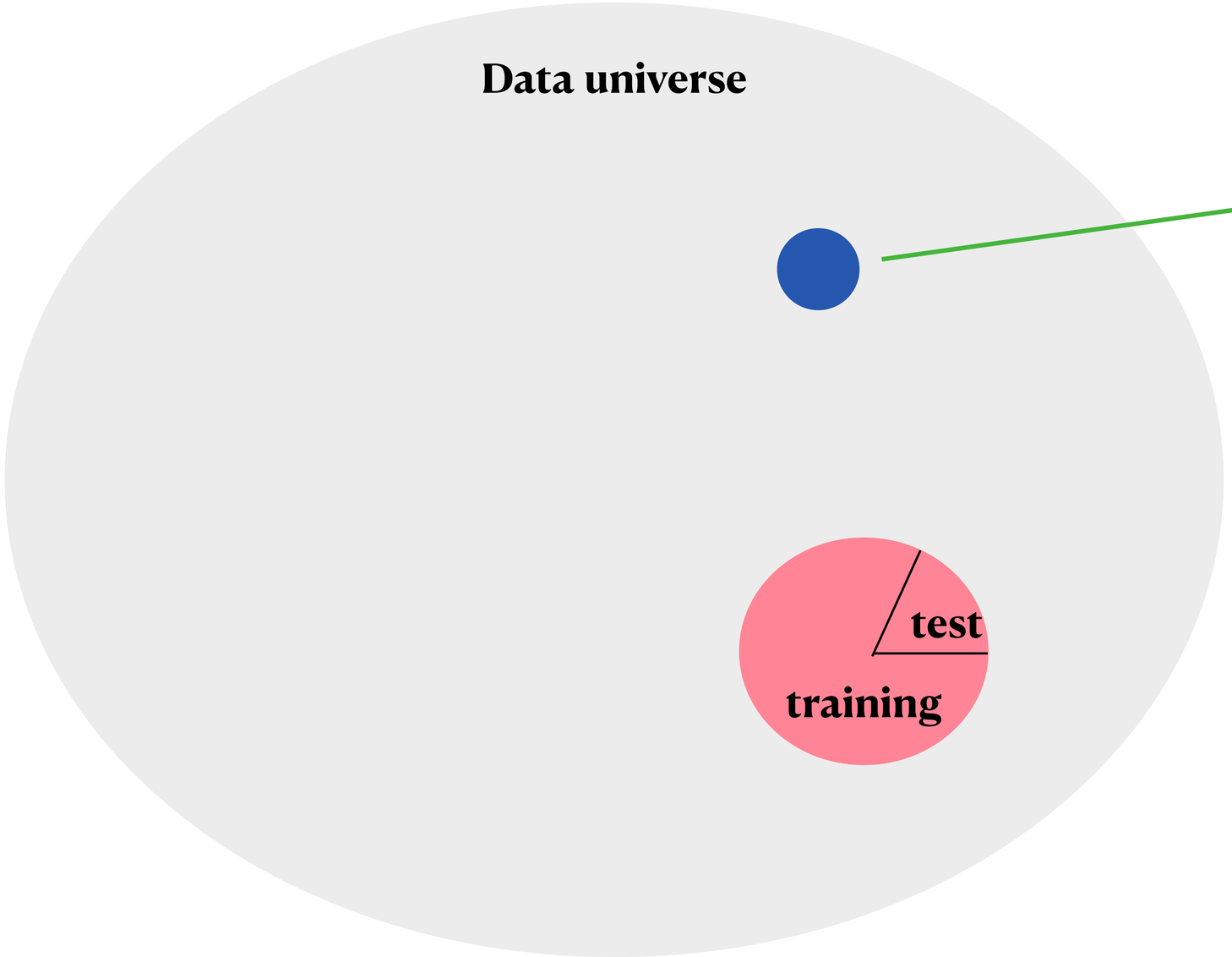
Ultimate goal of ML: deal with unseen data

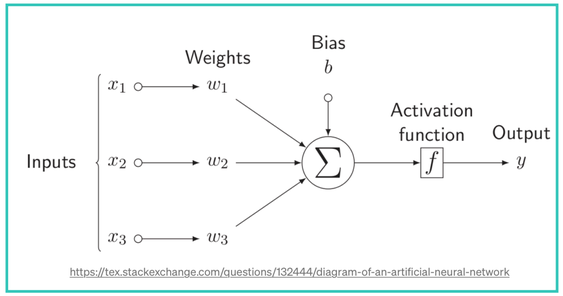




https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network

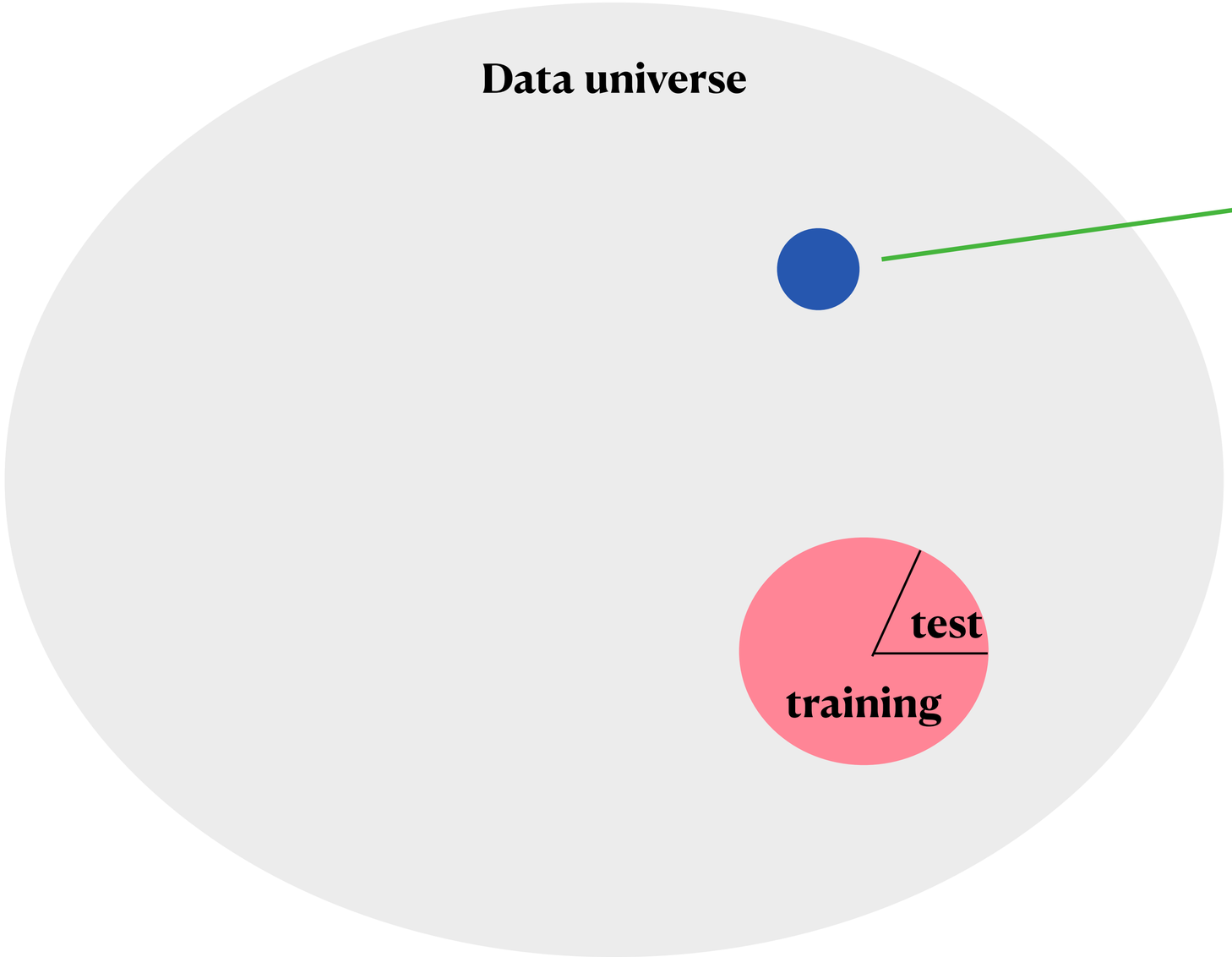
Ultimate goal of ML: deal with unseen data

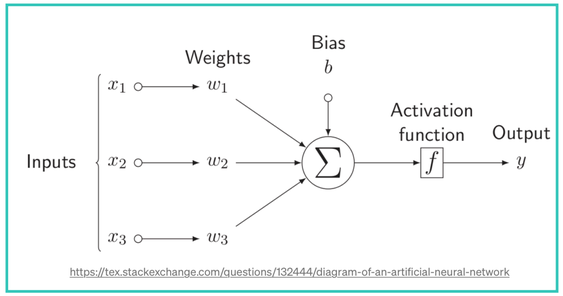




https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network

Ultimate goal of ML: deal with unseen data

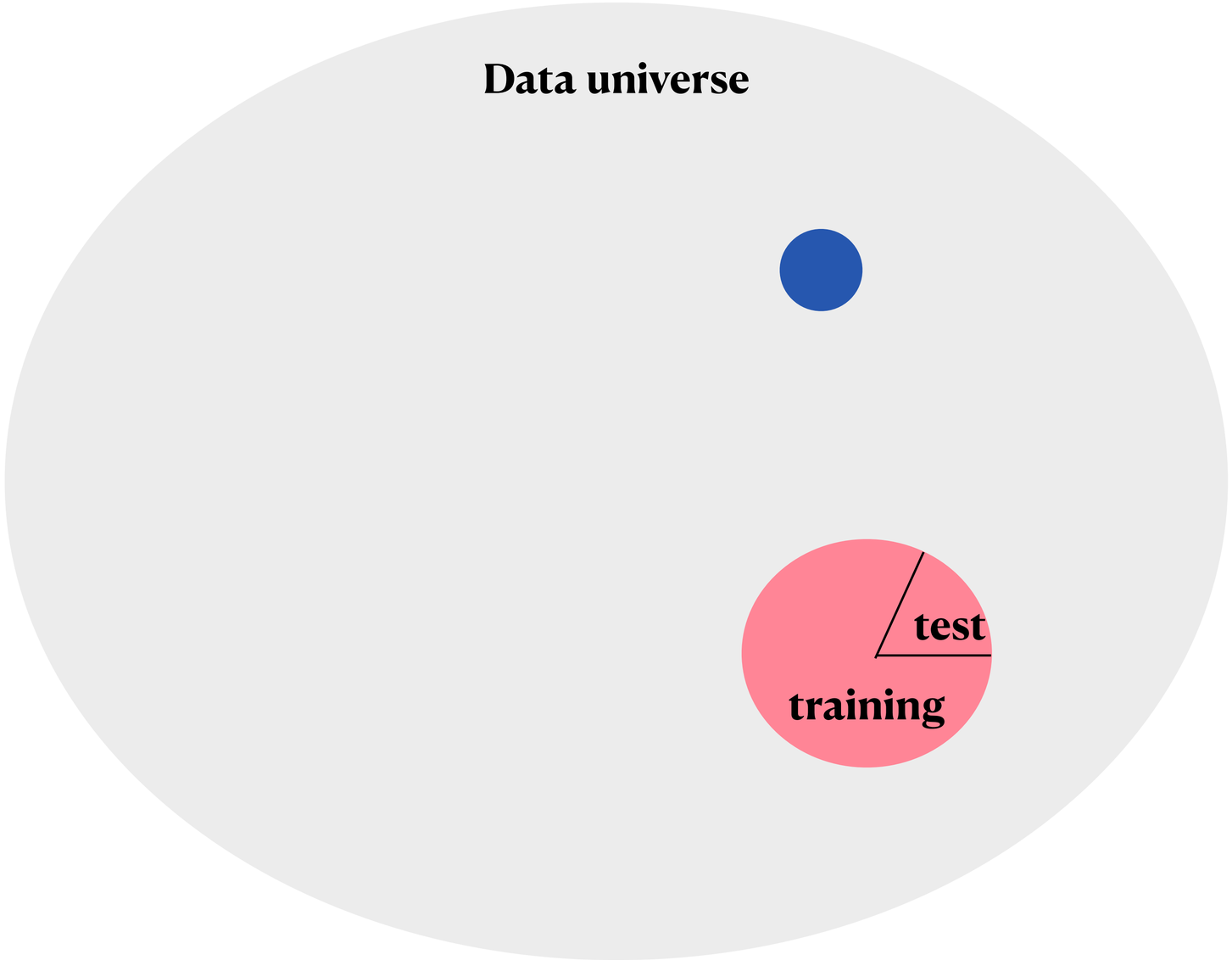




https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network

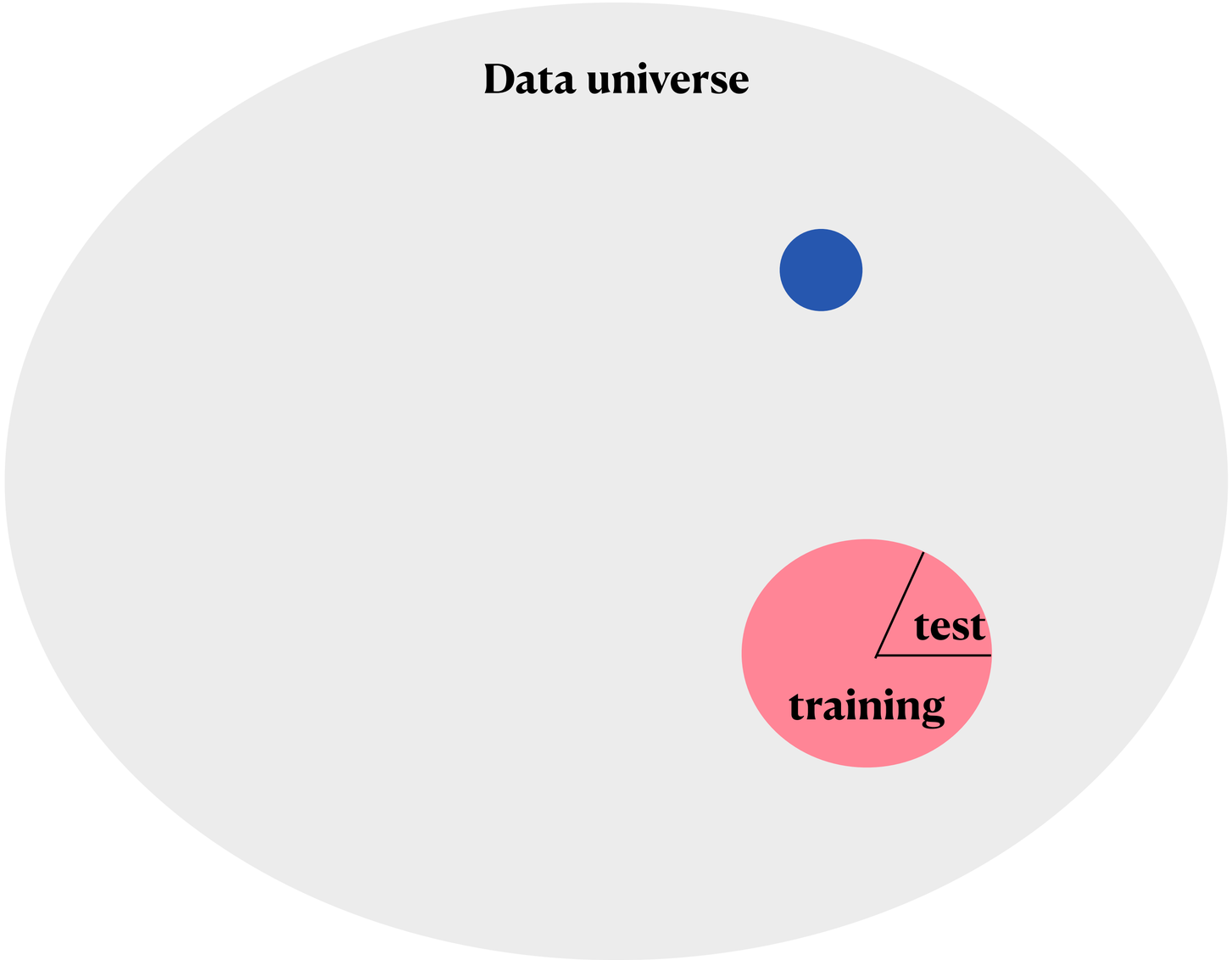
Golden question: what is the average error when our trained model applies to the unseen data?

Scenario 1



Data set: pink portion

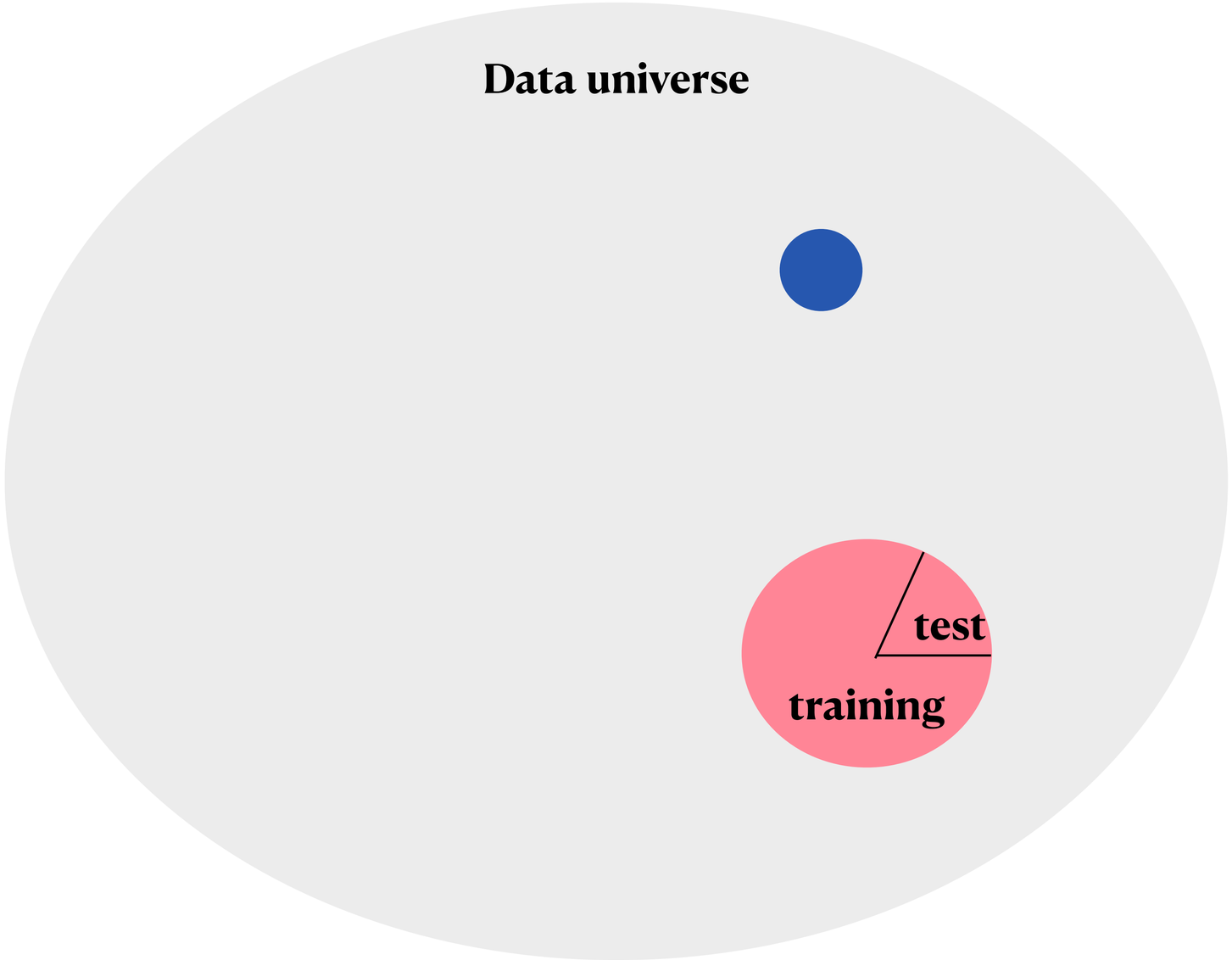
Scenario 1



Data set: pink portion

Split: 20% as test, 80% as training

Scenario 1

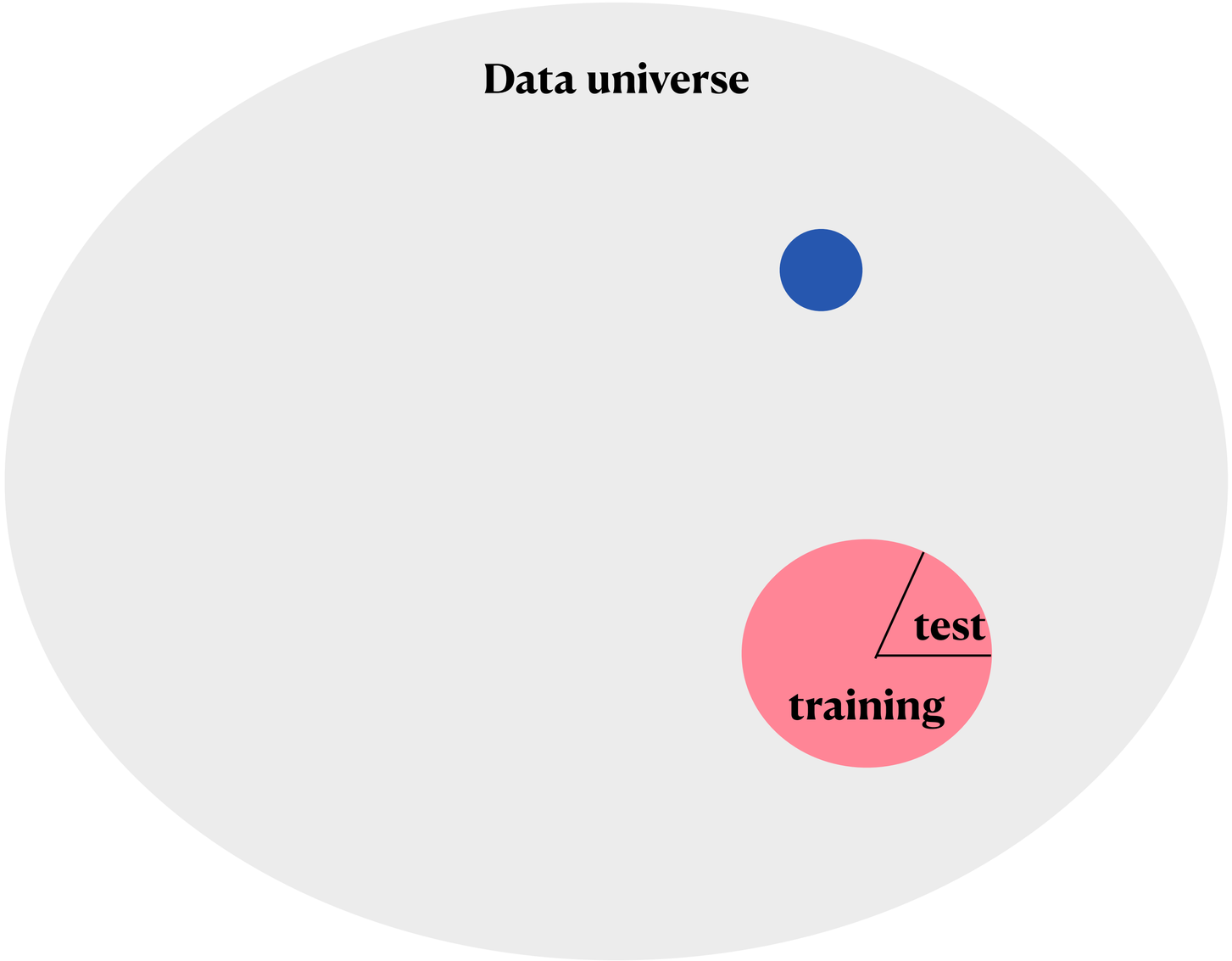


Data set: pink portion

Split: 20% as test, 80% as training

Specify model, loss function

Scenario 1



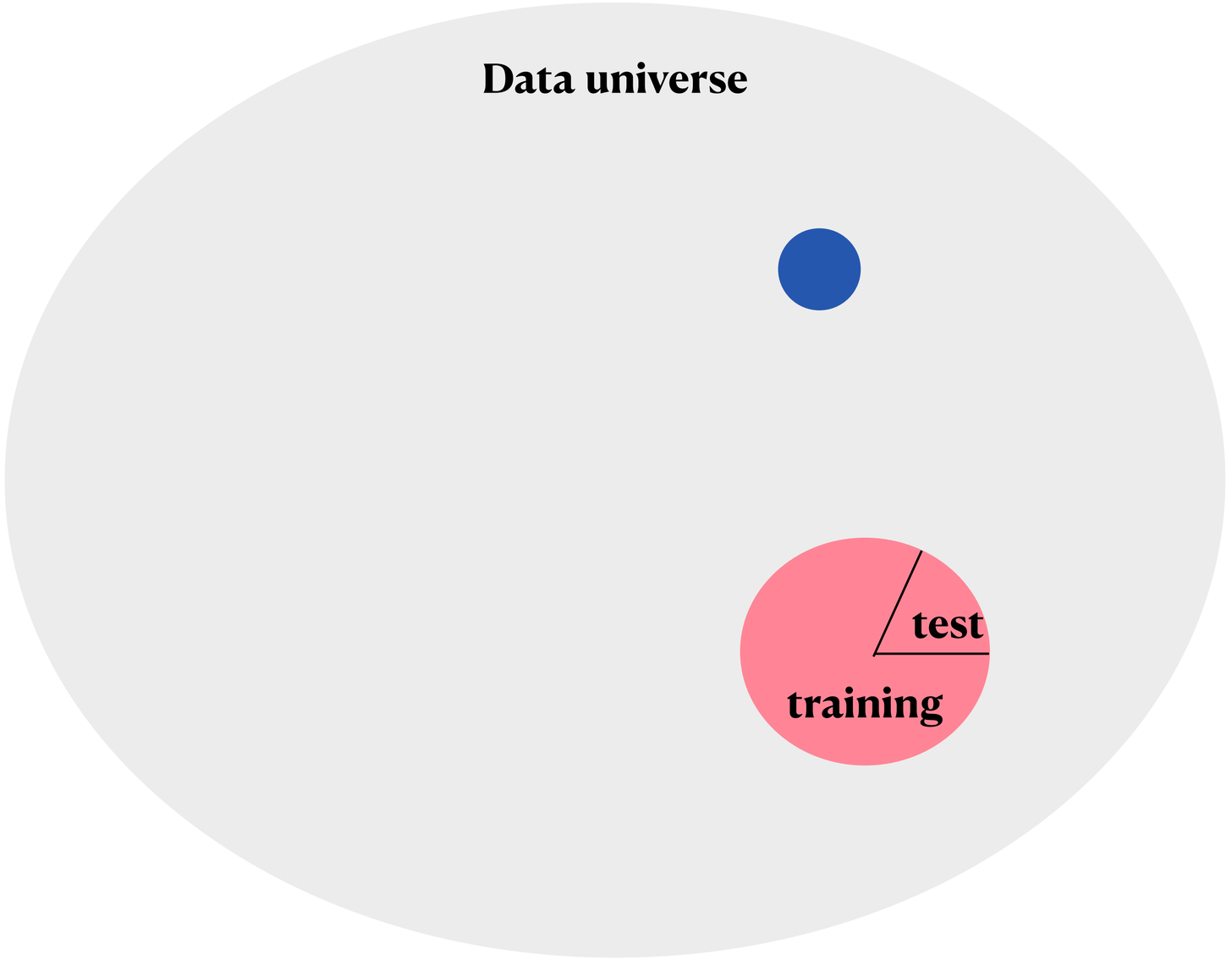
Data set: pink portion

Split: 20% as test, 80% as training

Specify model, loss function

Fit the model with training data —> get as small training error as you can

Scenario 1



Data set: pink portion

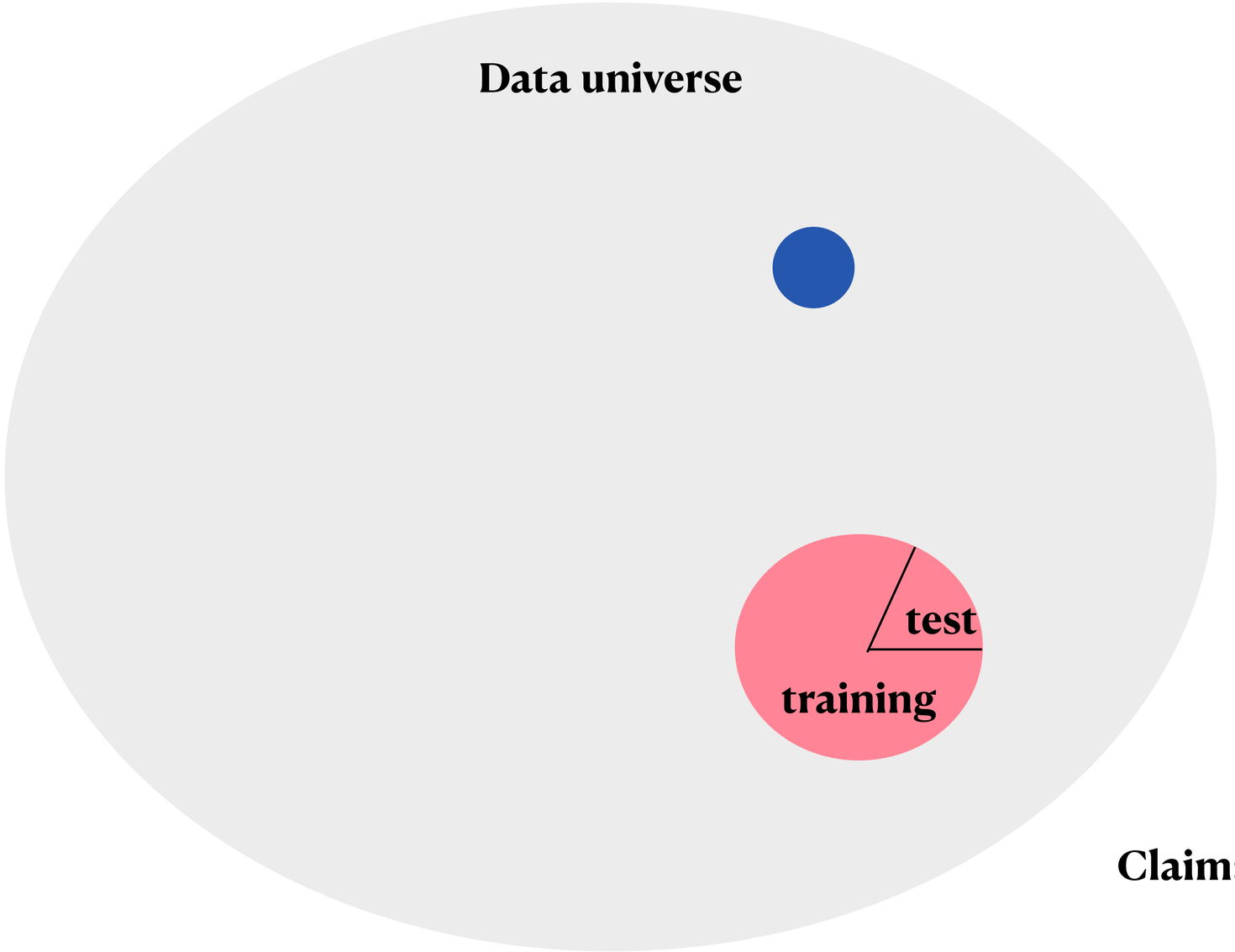
Split: 20% as test, 80% as training

Specify model, loss function

Fit the model with training data —> get as small training error as you can

Send test data into model —> get test error

Scenario 1



Data set: pink portion

Split: 20% as test, 80% as training

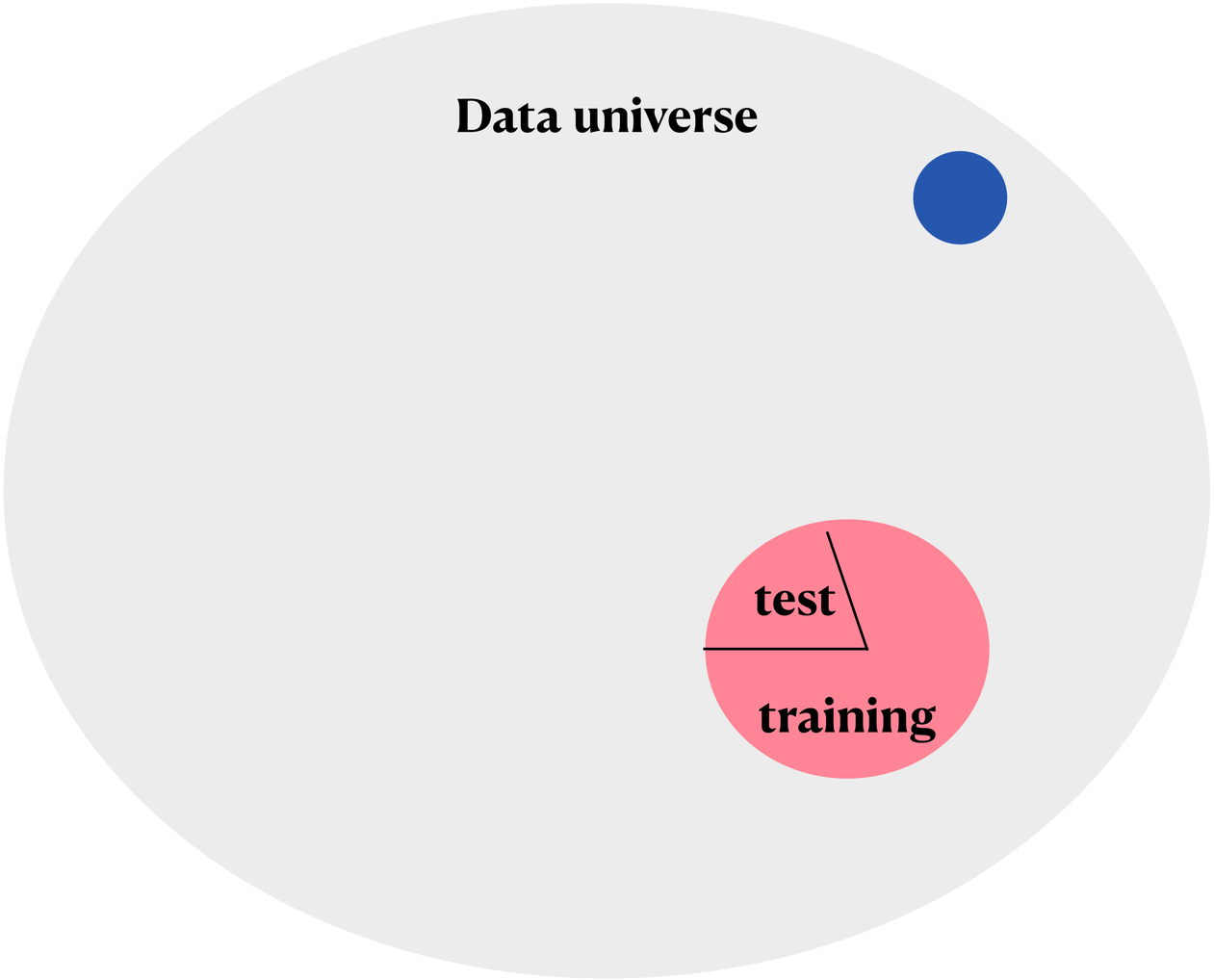
Specify model, loss function

Fit the model with training data —> get as small training error as you can

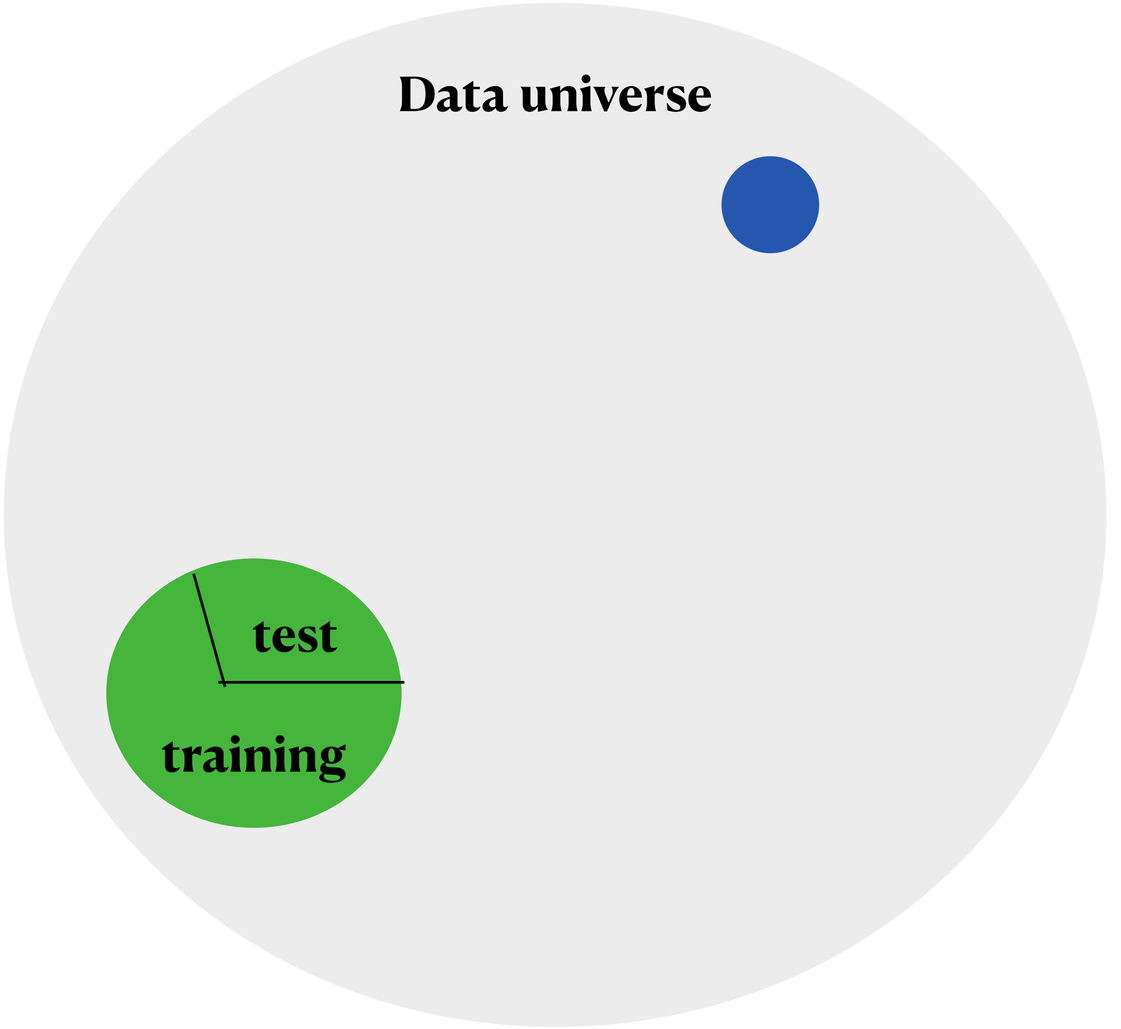
Send test data into model —> get test error

Claim: the average error on unseen data = test error

Team 1



Team 2



Very important concept in ML

	Training subset 1	**Training subset 2**	**Training subset 3**
Model 1 $f(x) = ax + b$	Estimated test error		
Model 2 $f(x) = ax + bx^2 + c$			
Model 3 $f(x) = ax + bx^2 + cx^3 + d$			

Very important concept in ML

	Training subset 1	**Training subset 2**	**Training subset 3**
Model 1 $f(x) = ax + b$	Estimated test error		
Model 2 $f(x) = ax + bx^2 + c$			
Model 3 $f(x) = ax + bx^2 + cx^3 + d$			

Very important concept in ML

	Training subset 1	**Training subset 2**	**Training subset 3**
Model 1 $f(x) = ax + b$	Estimated test error		
Model 2 $f(x) = ax + bx^2 + c$			
Model 3 $f(x) = ax + bx^2 + cx^3 + d$			

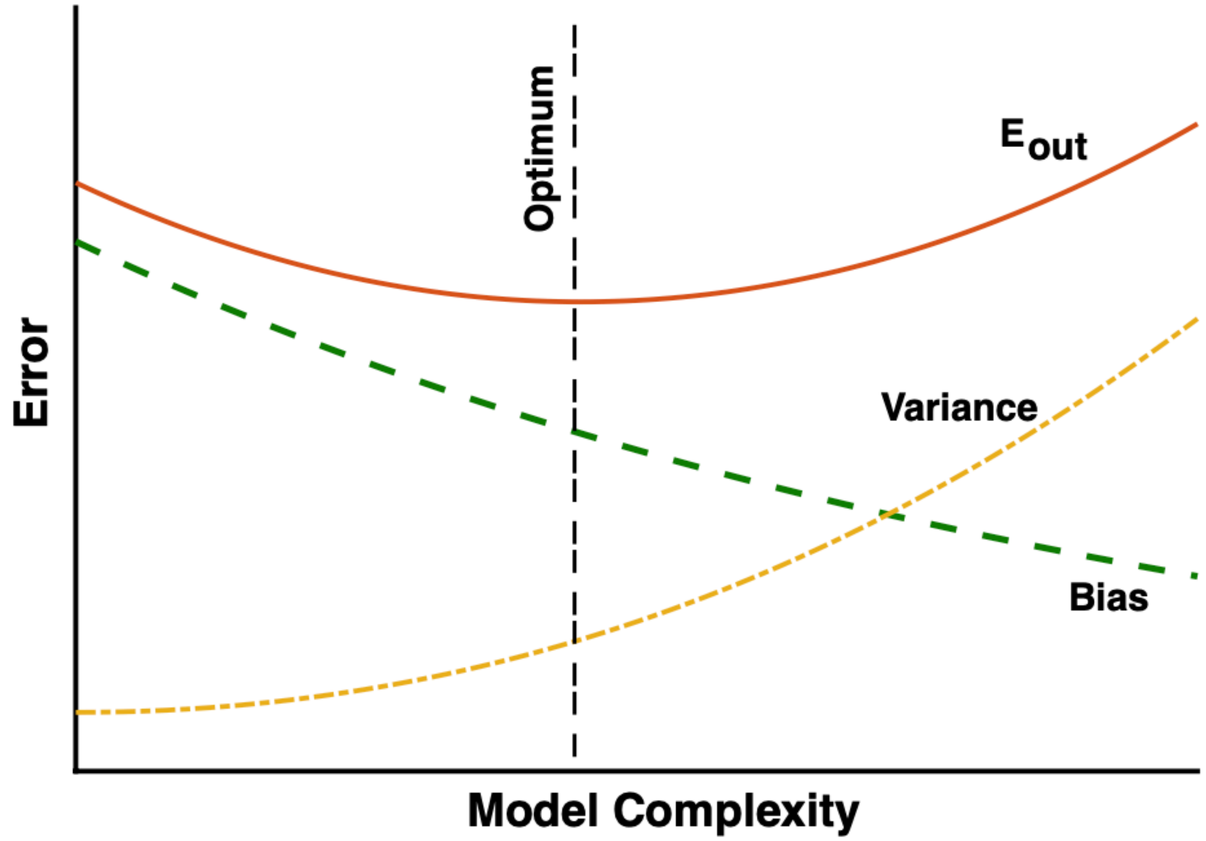
Mean —> bias
Variance —> variance

Very important concept in ML

	Training subset 1	**Training subset 2**	**Training subset 3**
Model 1 $f(x) = ax + b$	Estimated test error		
Model 2 $f(x) = ax + bx^2 + c$			
Model 3 $f(x) = ax + bx^2 + cx^3 + d$			

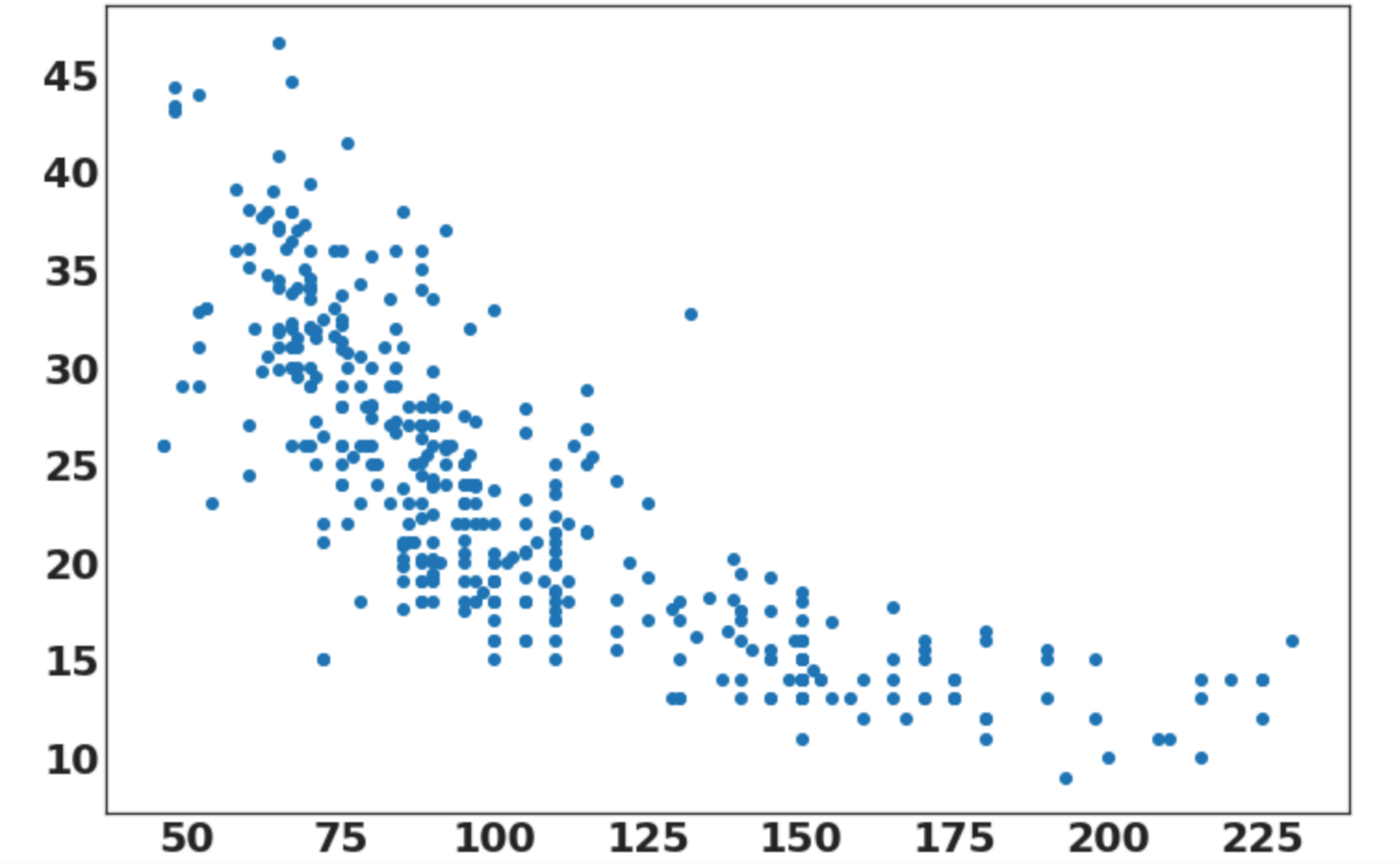
Mean —> bias
Variance —> variance

Bias-variance tradeoff



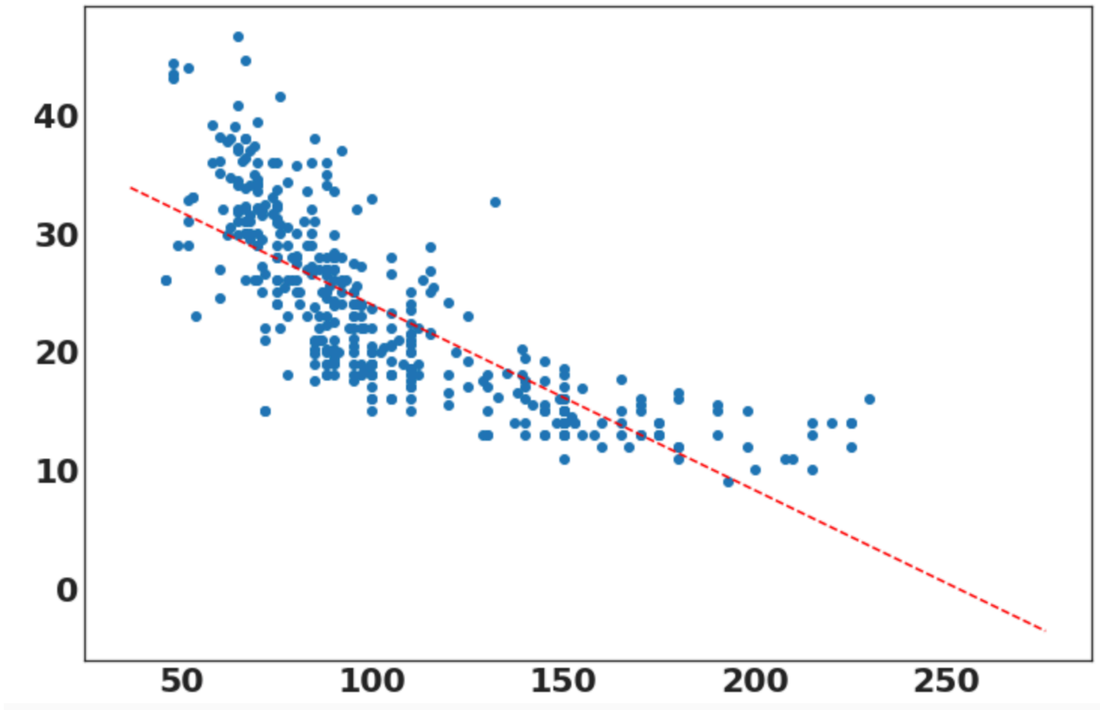
Auto data set

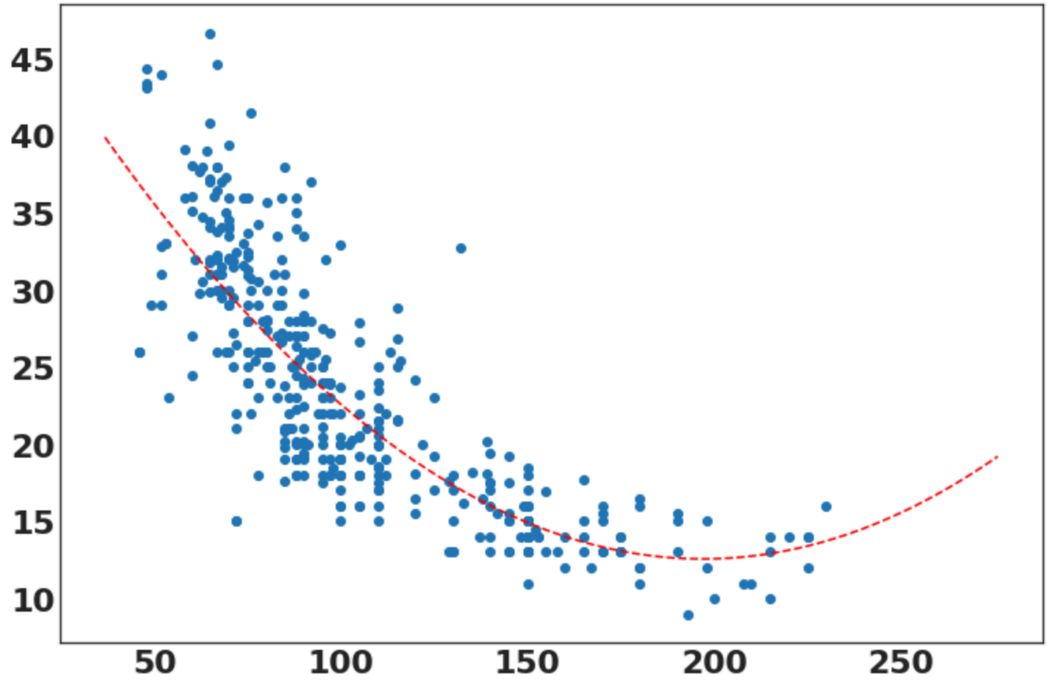
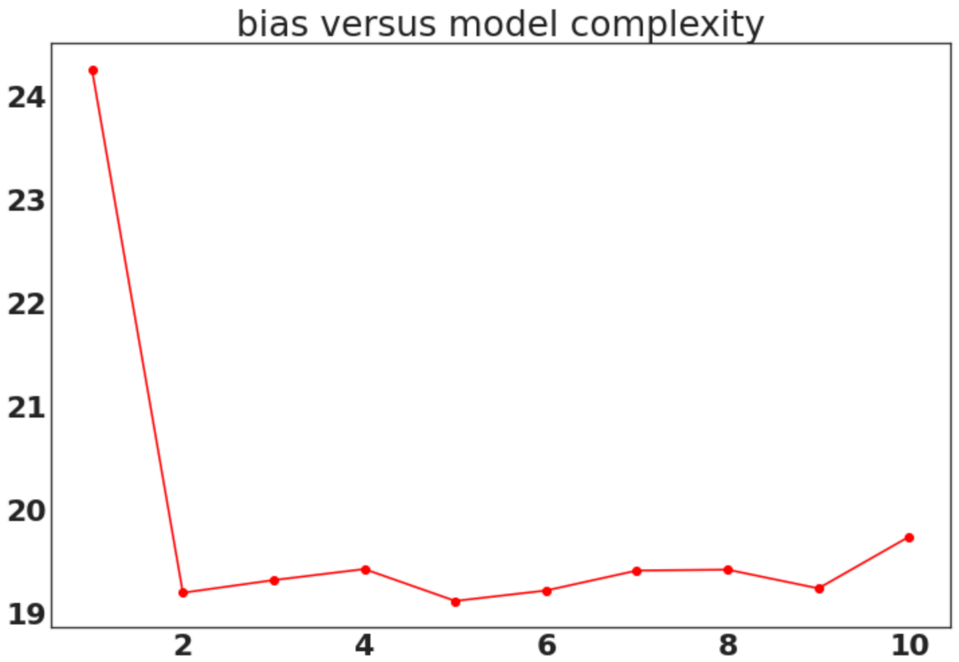
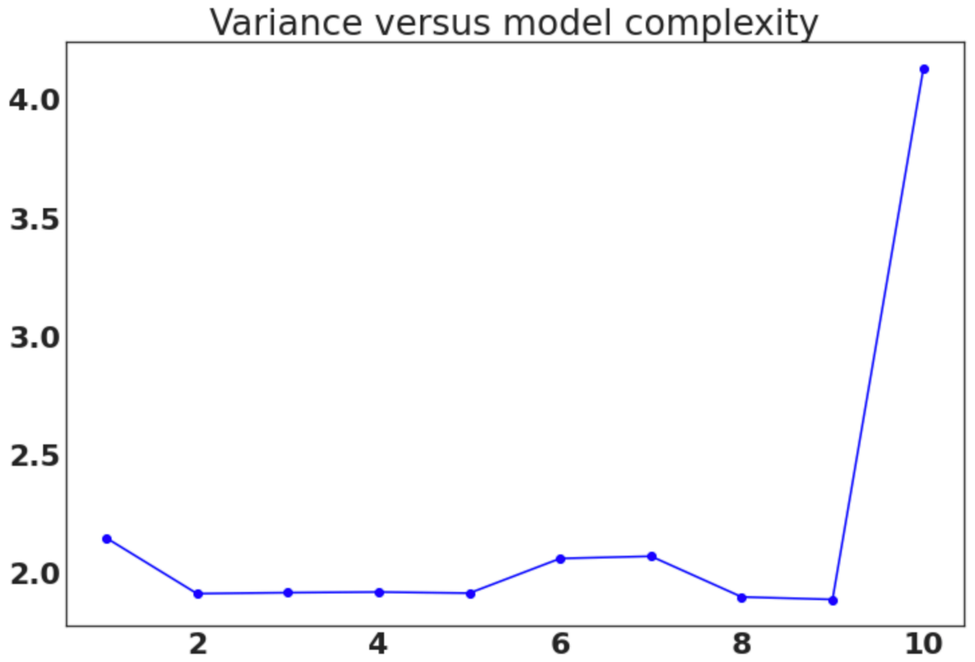
Mile per gallon



Horsepower

Linear versus quadratic models





Bias-variance tradeoff
(See Lab code for simulation)