Statistics and Machine Learning

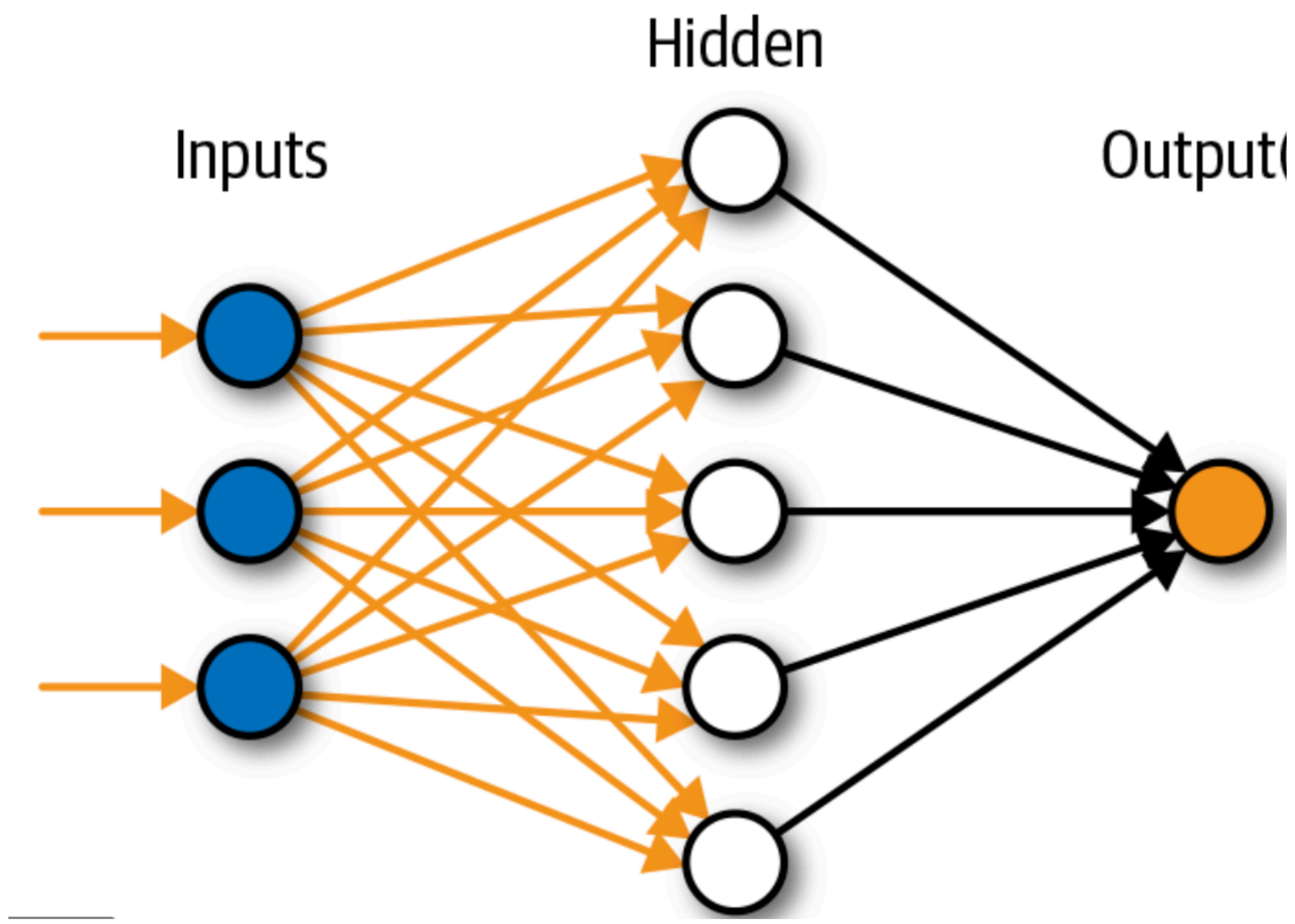
Week 2
01/25 — 01/29
Introduction to Machine Learning

Contents for week 2

- Supervised learning
- Example: digit classification
- Elements of supervised learning
- Why statistics
- Unsupervised learning
- Clustering, principal component analysis, SVM
- Recommender system, natural language processing
- Lab: basic python

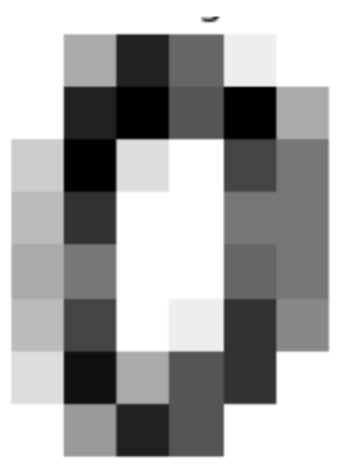
Artificial neuron does the job

Google the keyword "pytorch", "Keras", "tensor flow" to learn more



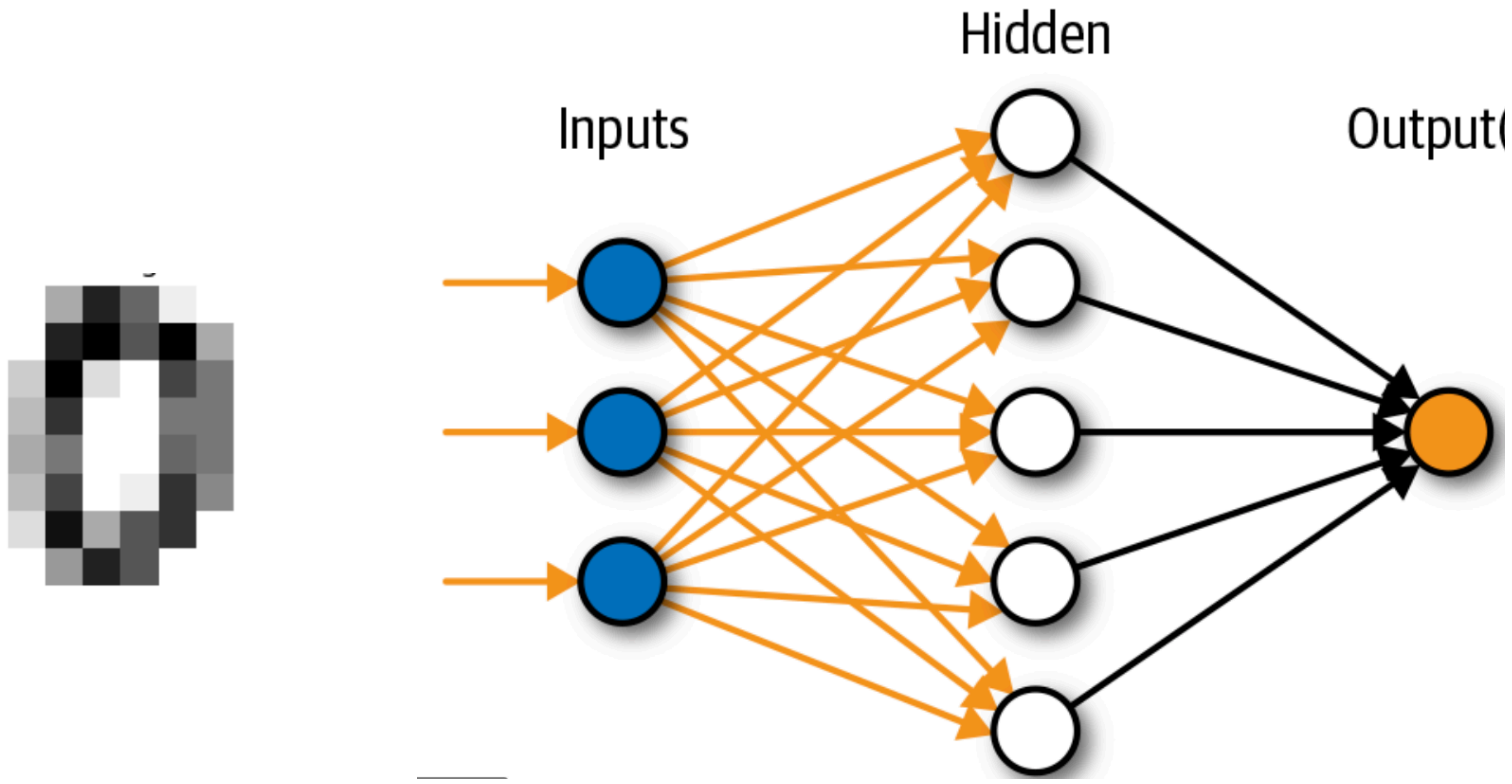
A horse with 99% confidence

Hand-written digit images



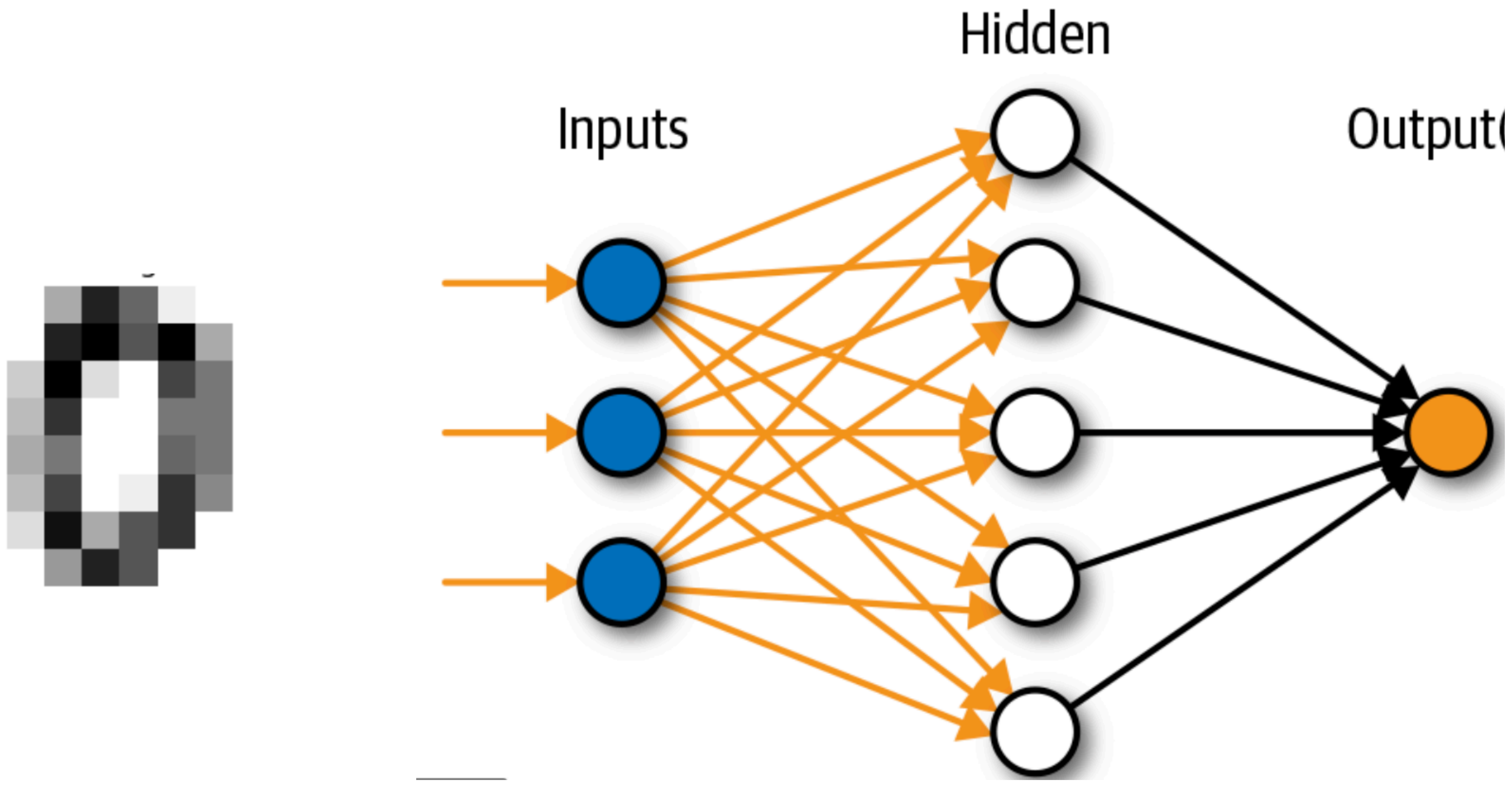
https://scikit-learn.org/stable/auto_examples/classification/plot_digits_classification.html

Hand-written digit images



https://scikit-learn.org/stable/auto_examples/classification/plot_digits_classification.html

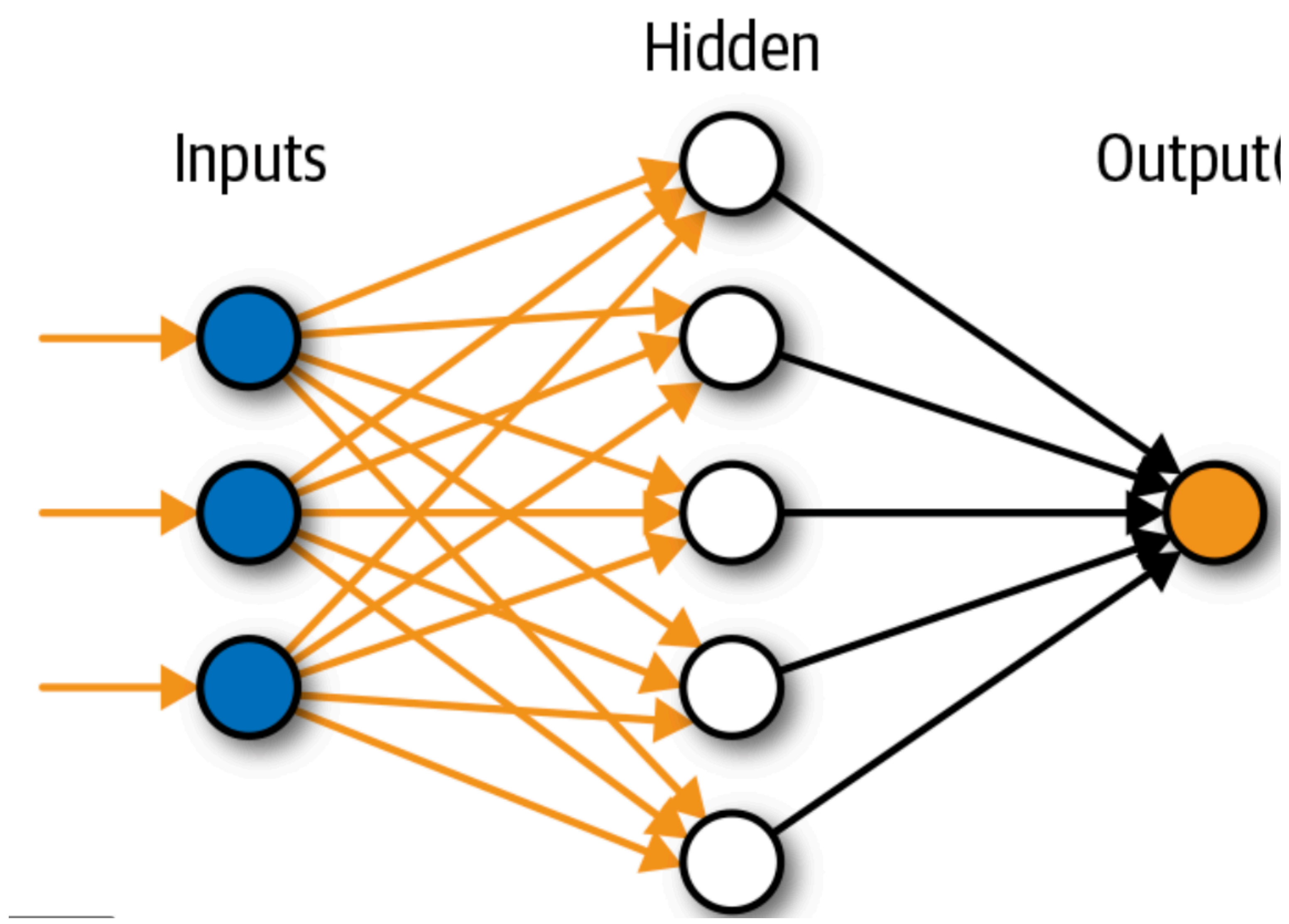
Hand-written digit images



It is o

https://scikit-learn.org/stable/auto_examples/classification/plot_digits_classification.html

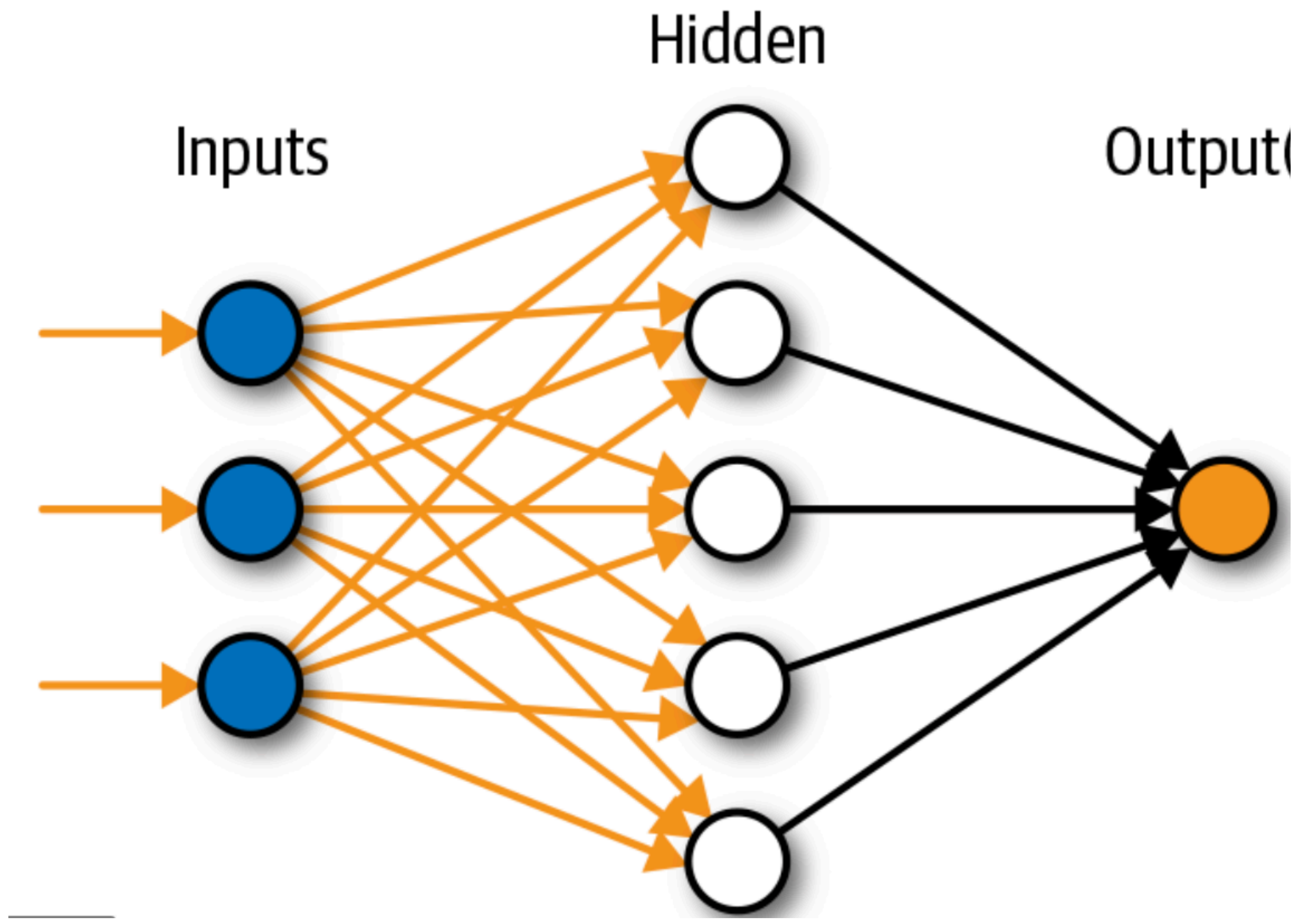
Hand-written digit images



It is o

https://scikit-learn.org/stable/auto_examples/classification/plot_digits_classification.html

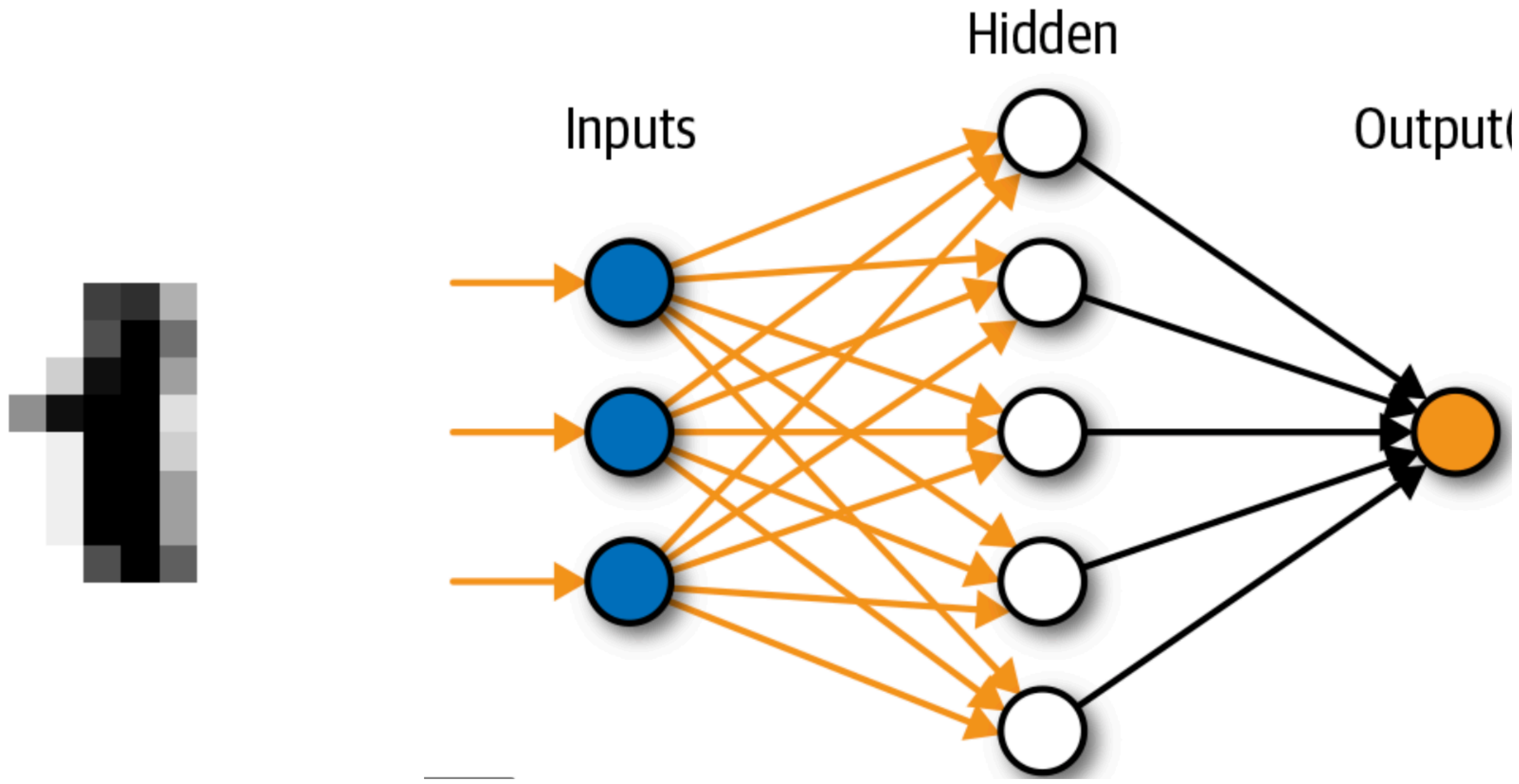
Hand-written digit images



It is o

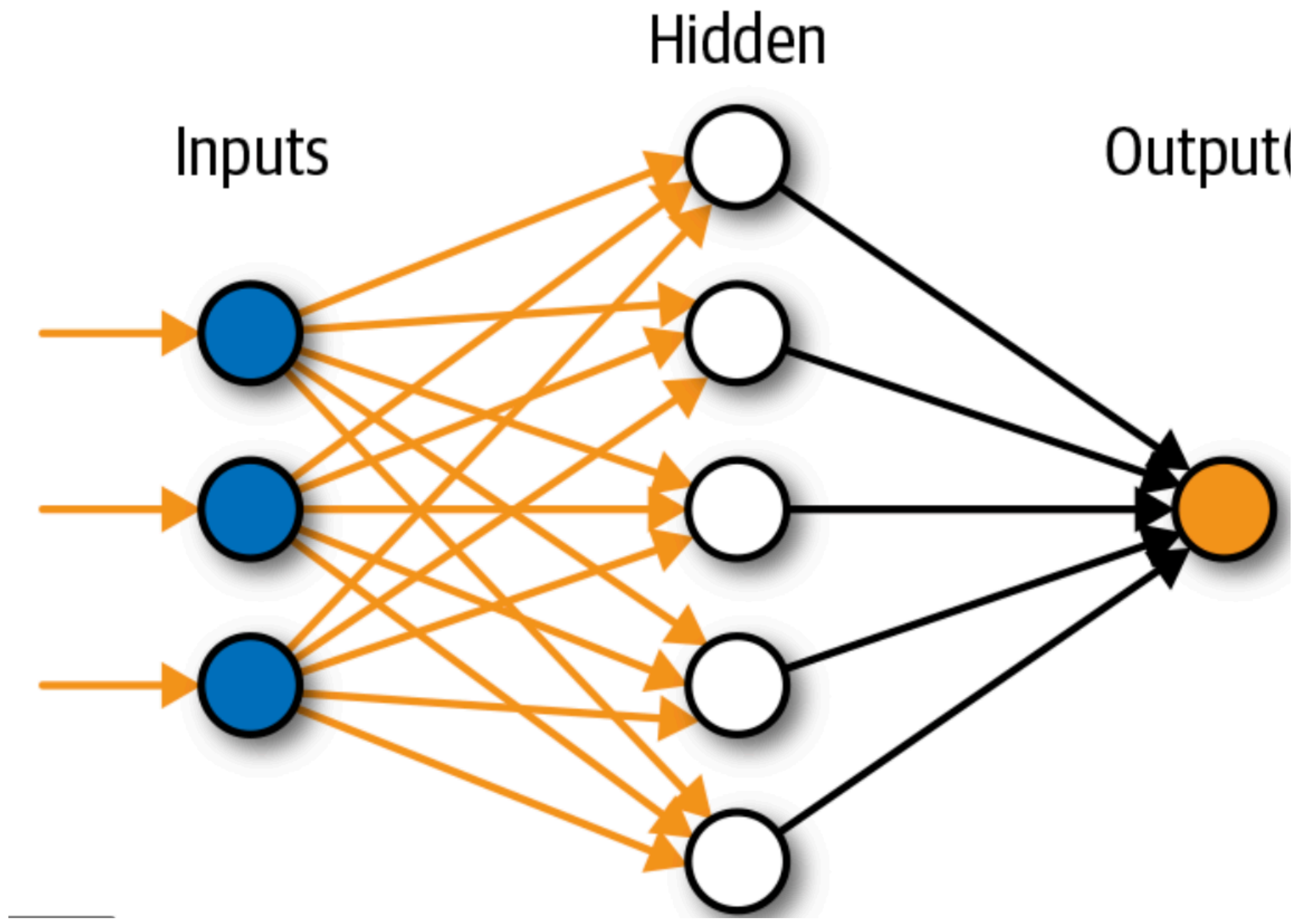
https://scikit-learn.org/stable/auto_examples/classification/plot_digits_classification.html

Hand-written digit images



https://scikit-learn.org/stable/auto_examples/classification/plot_digits_classification.html

Hand-written digit images



It is 1

https://scikit-learn.org/stable/auto_examples/classification/plot_digits_classification.html

Clarification on datum versus data

Singular versus plural

datum noun

Save Word

da·tum | \ ˈdā-təm , ˈda- , ˈdä- \

Definition of *datum*

1 *plural* **data**\ ˈdā-tə , ˈda- *also* ˈdä- \ **:** something given or admitted especially as a basis for reasoning or inference

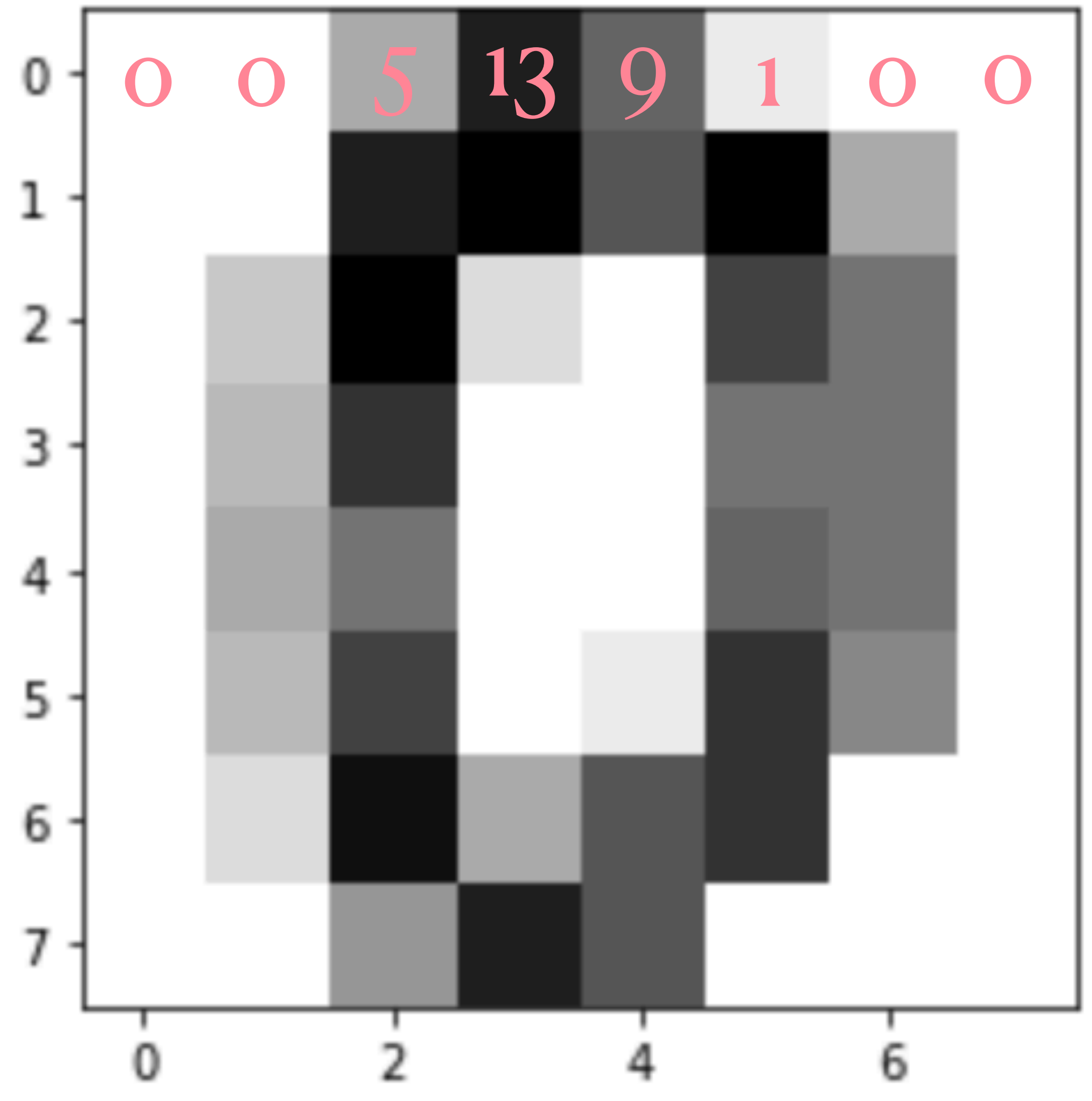
// an important historical *datum*

// This enormous expense—and considerable risk—to pick up a *datum* or two about geriatrics?

— Charles Krauthammer

Model, data set, and loss function

Elements in machine learning




```
digits.images[0]
```

```
array([[ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.],
       [ 0.,  0., 13., 15., 10., 15.,  5.,  0.],
       [ 0.,  3., 15.,  2.,  0., 11.,  8.,  0.],
       [ 0.,  4., 12.,  0.,  0.,  8.,  8.,  0.],
       [ 0.,  5.,  8.,  0.,  0.,  9.,  8.,  0.],
       [ 0.,  4., 11.,  0.,  1., 12.,  7.,  0.],
       [ 0.,  2., 14.,  5., 10., 12.,  0.,  0.],
       [ 0.,  0.,  6., 13., 10.,  0.,  0.,  0.]])
```

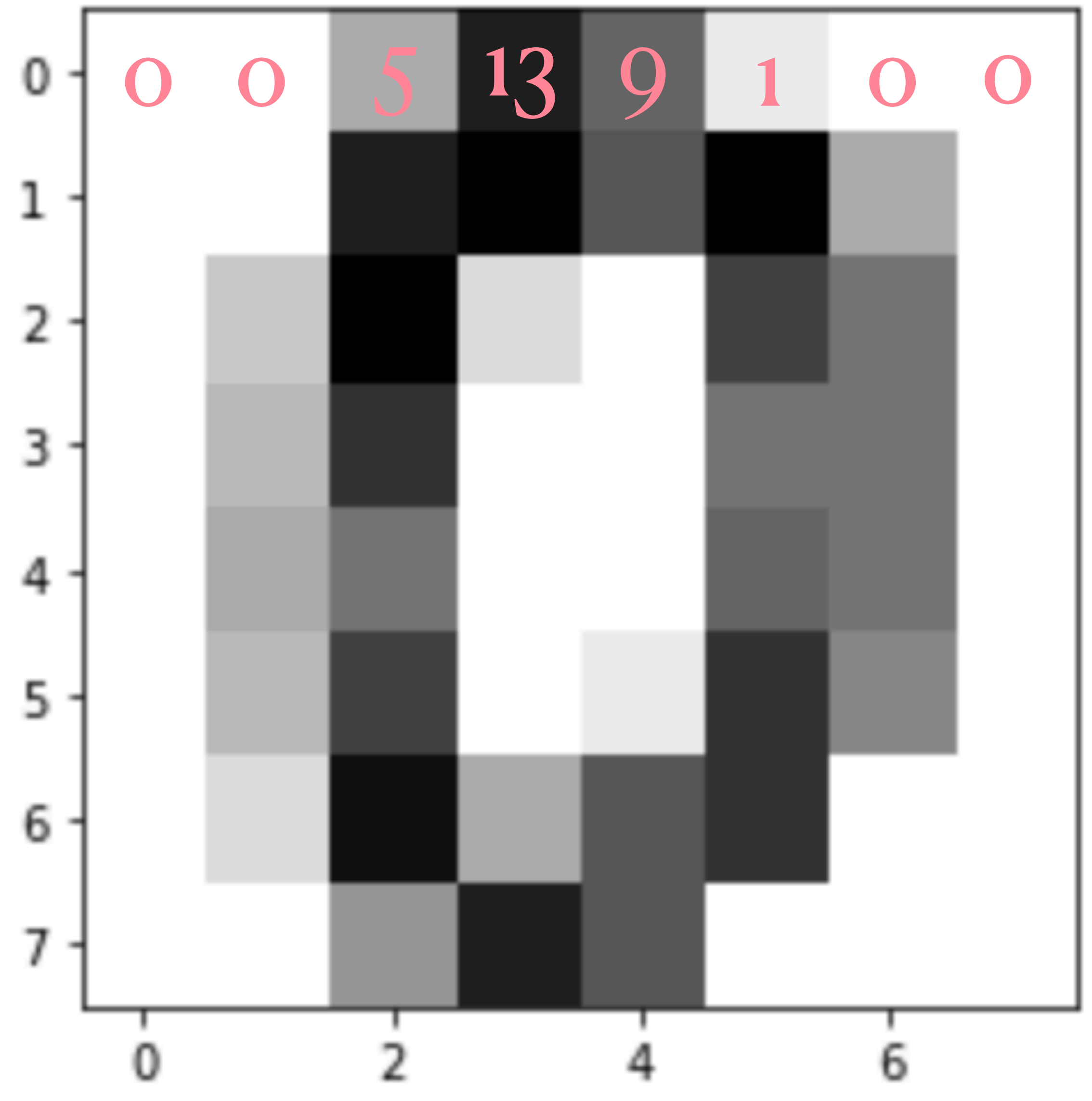
```
# target_names store the label for the digit image

digits.target_names[0]
```

```
0
```

See the code in MNIST_basic.ipynb

Model, data set, and loss function

Elements in machine learning




```
digits.images[0]
```

```
array([[ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.],
       [ 0.,  0., 13., 15., 10., 15.,  5.,  0.],
       [ 0.,  3., 15.,  2.,  0., 11.,  8.,  0.],
       [ 0.,  4., 12.,  0.,  0.,  8.,  8.,  0.],
       [ 0.,  5.,  8.,  0.,  0.,  9.,  8.,  0.],
       [ 0.,  4., 11.,  0.,  1., 12.,  7.,  0.],
       [ 0.,  2., 14.,  5., 10., 12.,  0.,  0.],
       [ 0.,  0.,  6., 13., 10.,  0.,  0.,  0.]])
```

```
# target_names store the label for the digit image

digits.target_names[0]
```

```
0
```

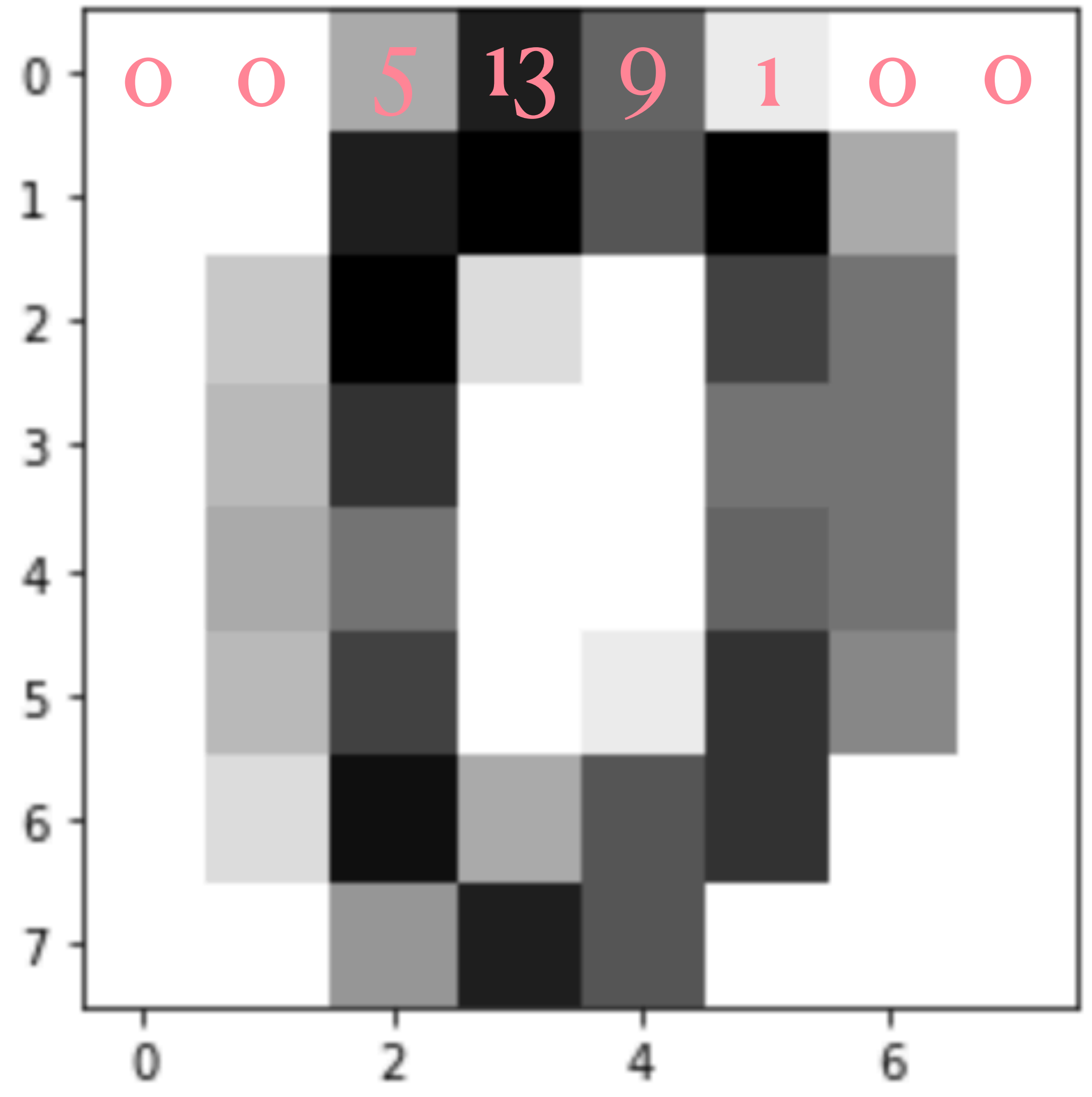
The whole thing in the left is a datum. The information contained:

1. Image (8x8, 64 numbers)
2. Label '0'

See the code in MNIST_basic.ipynb

Model, data set, and loss function

Elements in machine learning




```
digits.images[0]
```

```
array([[ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.],
       [ 0.,  0., 13., 15., 10., 15.,  5.,  0.],
       [ 0.,  3., 15.,  2.,  0., 11.,  8.,  0.],
       [ 0.,  4., 12.,  0.,  0.,  8.,  8.,  0.],
       [ 0.,  5.,  8.,  0.,  0.,  9.,  8.,  0.],
       [ 0.,  4., 11.,  0.,  1., 12.,  7.,  0.],
       [ 0.,  2., 14.,  5., 10., 12.,  0.,  0.],
       [ 0.,  0.,  6., 13., 10.,  0.,  0.,  0.]])
```

```
# target_names store the label for the digit image

digits.target_names[0]
```

```
0
```

The whole thing in the left is a datum. The information contained:

1. Image (8x8, 64 numbers)
2. Label '0'

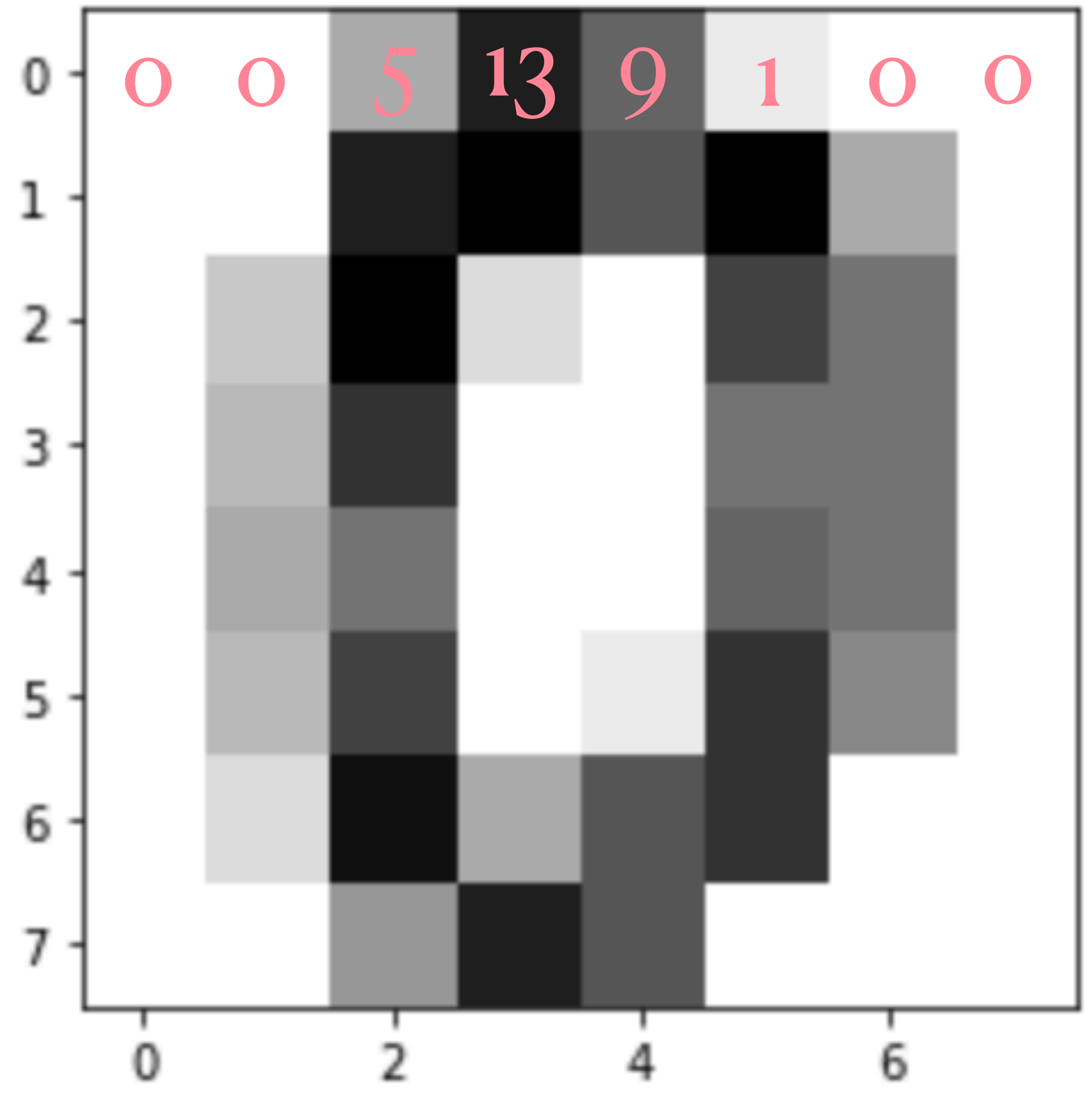
For convenience, we usually do

1. Reshape the image into a 64x1 vector, call it x
2. Call y = 0

See the code in MNIST_basic.ipynb

Model, data set, and loss function

Elements in machine learning




```
digits.images[0]
```

```
array([[ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.],
       [ 0.,  0., 13., 15., 10., 15.,  5.,  0.],
       [ 0.,  3., 15.,  2.,  0., 11.,  8.,  0.],
       [ 0.,  4., 12.,  0.,  0.,  8.,  8.,  0.],
       [ 0.,  5.,  8.,  0.,  0.,  9.,  8.,  0.],
       [ 0.,  4., 11.,  0.,  1., 12.,  7.,  0.],
       [ 0.,  2., 14.,  5., 10., 12.,  0.,  0.],
       [ 0.,  0.,  6., 13., 10.,  0.,  0.,  0.]])
```

< — x

The whole thing in the left is a datum. The information contained:

1. Image (8x8, 64 numbers)
2. Label ‘0’

For convenience, we usually do

1. Reshape the image into a 64x1 vector, call it x
2. Call y = 0

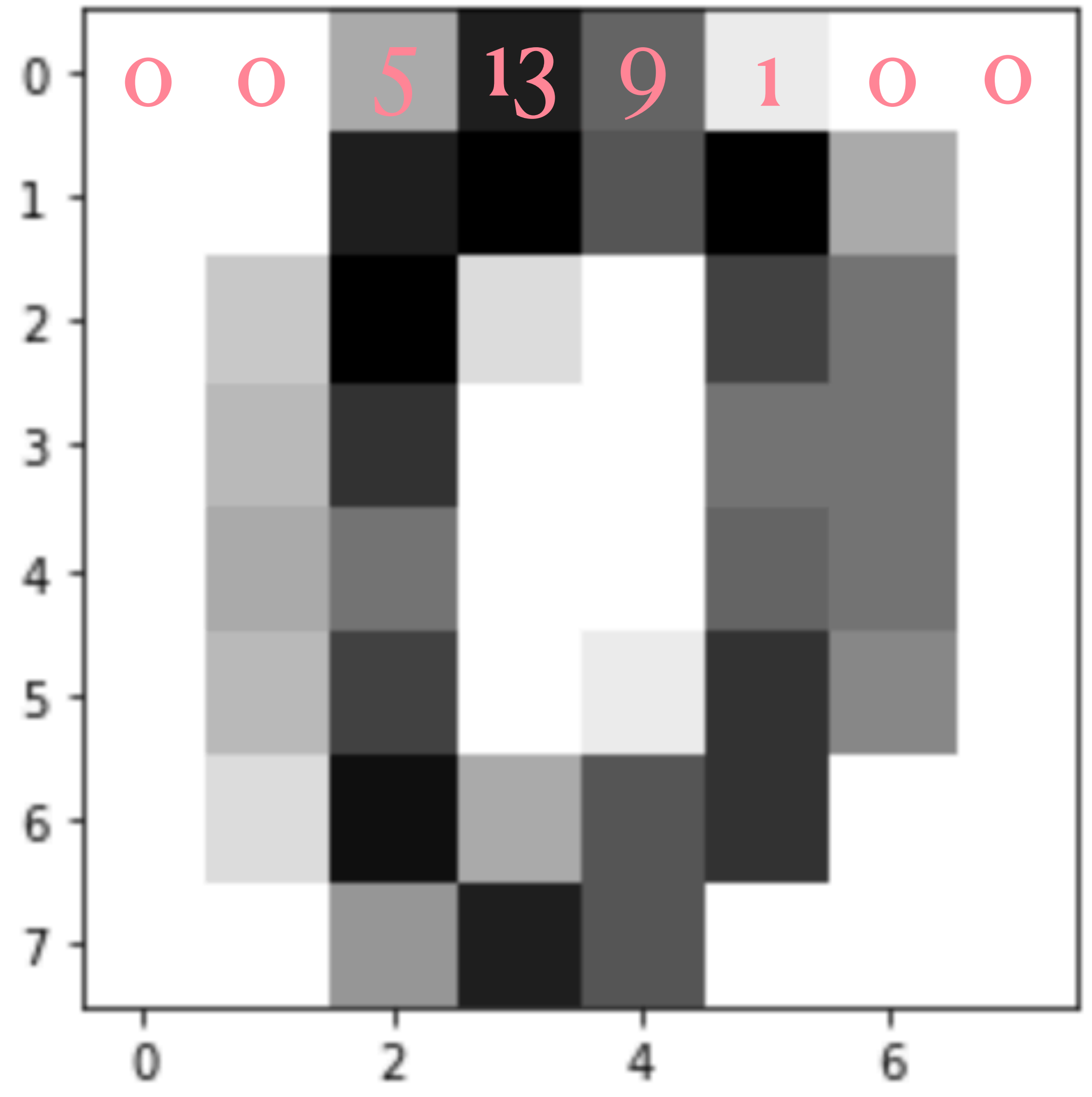
```
# target_names store the label for the digit image

digits.target_names[0]
```

```
0
```

See the code in MNIST_basic.ipynb

Model, data set, and loss function

Elements in machine learning




```
digits.images[0]
```

```
array([[ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.],
       [ 0.,  0., 13., 15., 10., 15.,  5.,  0.],
       [ 0.,  3., 15.,  2.,  0., 11.,  8.,  0.],
       [ 0.,  4., 12.,  0.,  0.,  8.,  8.,  0.],
       [ 0.,  5.,  8.,  0.,  0.,  9.,  8.,  0.],
       [ 0.,  4., 11.,  0.,  1., 12.,  7.,  0.],
       [ 0.,  2., 14.,  5., 10., 12.,  0.,  0.],
       [ 0.,  0.,  6., 13., 10.,  0.,  0.,  0.]])
```

< — x

```
# target_names store the label for the digit image

digits.target_names[0]
```

```
0
```

< — y

The whole thing in the left is a datum. The information contained:

1. Image (8x8, 64 numbers)
2. Label ‘0’

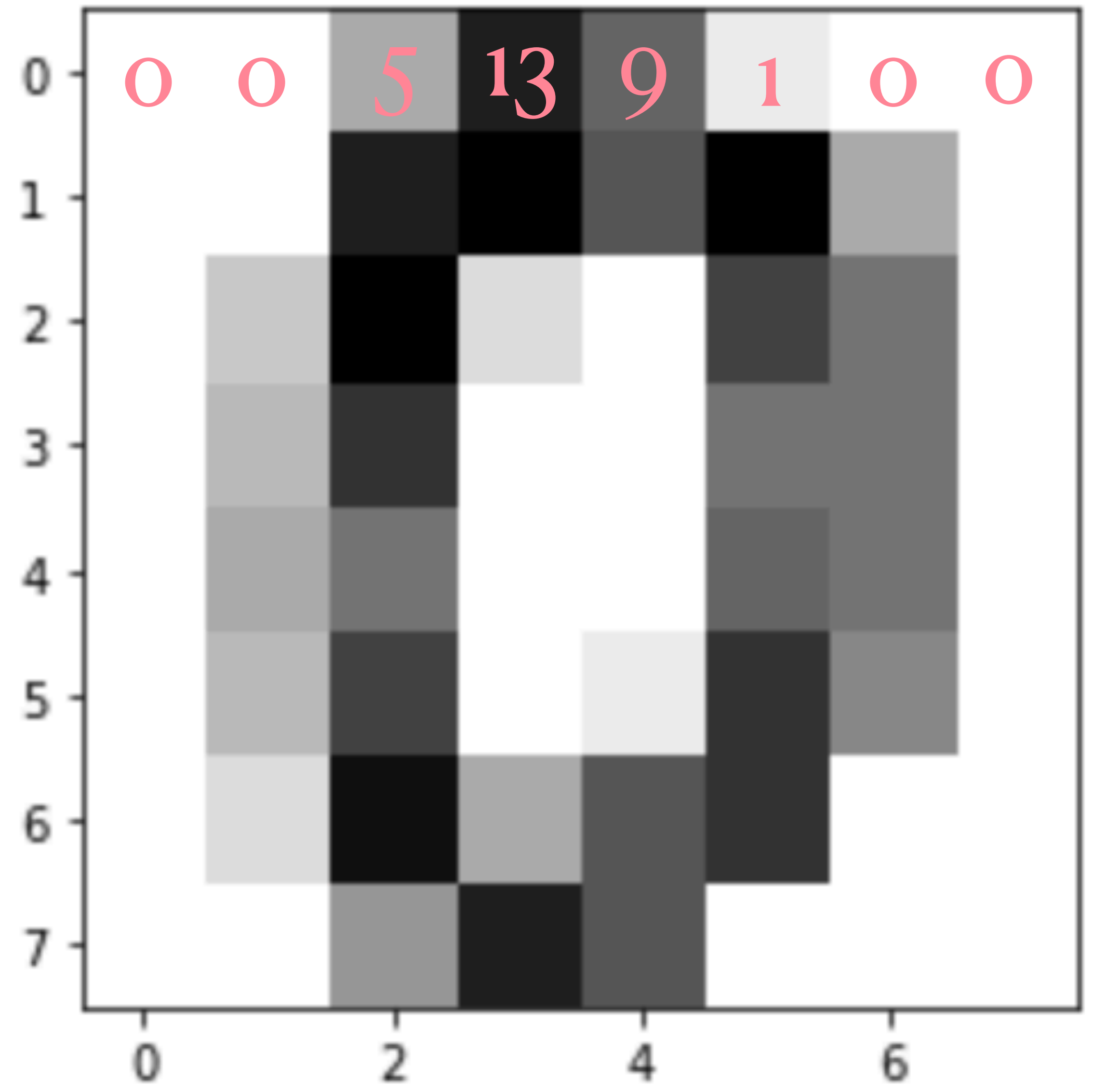
For convenience, we usually do

1. Reshape the image into a 64x1 vector, call it x
2. Call y = 0

See the code in MNIST_basic.ipynb

Model, data set, and loss function

Elements in machine learning




```
digits.images[0]
```

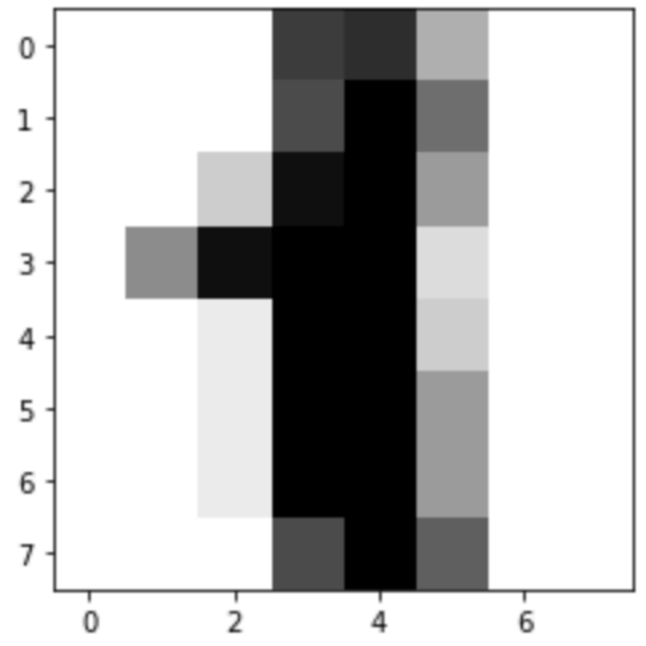
```
array([[ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.],
       [ 0.,  0., 13., 15., 10., 15.,  5.,  0.],
       [ 0.,  3., 15.,  2.,  0., 11.,  8.,  0.],
       [ 0.,  4., 12.,  0.,  0.,  8.,  8.,  0.],
       [ 0.,  5.,  8.,  0.,  0.,  9.,  8.,  0.],
       [ 0.,  4., 11.,  0.,  1., 12.,  7.,  0.],
       [ 0.,  2., 14.,  5., 10., 12.,  0.,  0.],
       [ 0.,  0.,  6., 13., 10.,  0.,  0.,  0.]])
```

< — x

```
# target_names store the label for the digit image

digits.target_names[0]
```

```
0
```

< — y



The whole thing in the left is a datum. The information contained:

1. Image (8x8, 64 numbers)
2. Label ‘0’

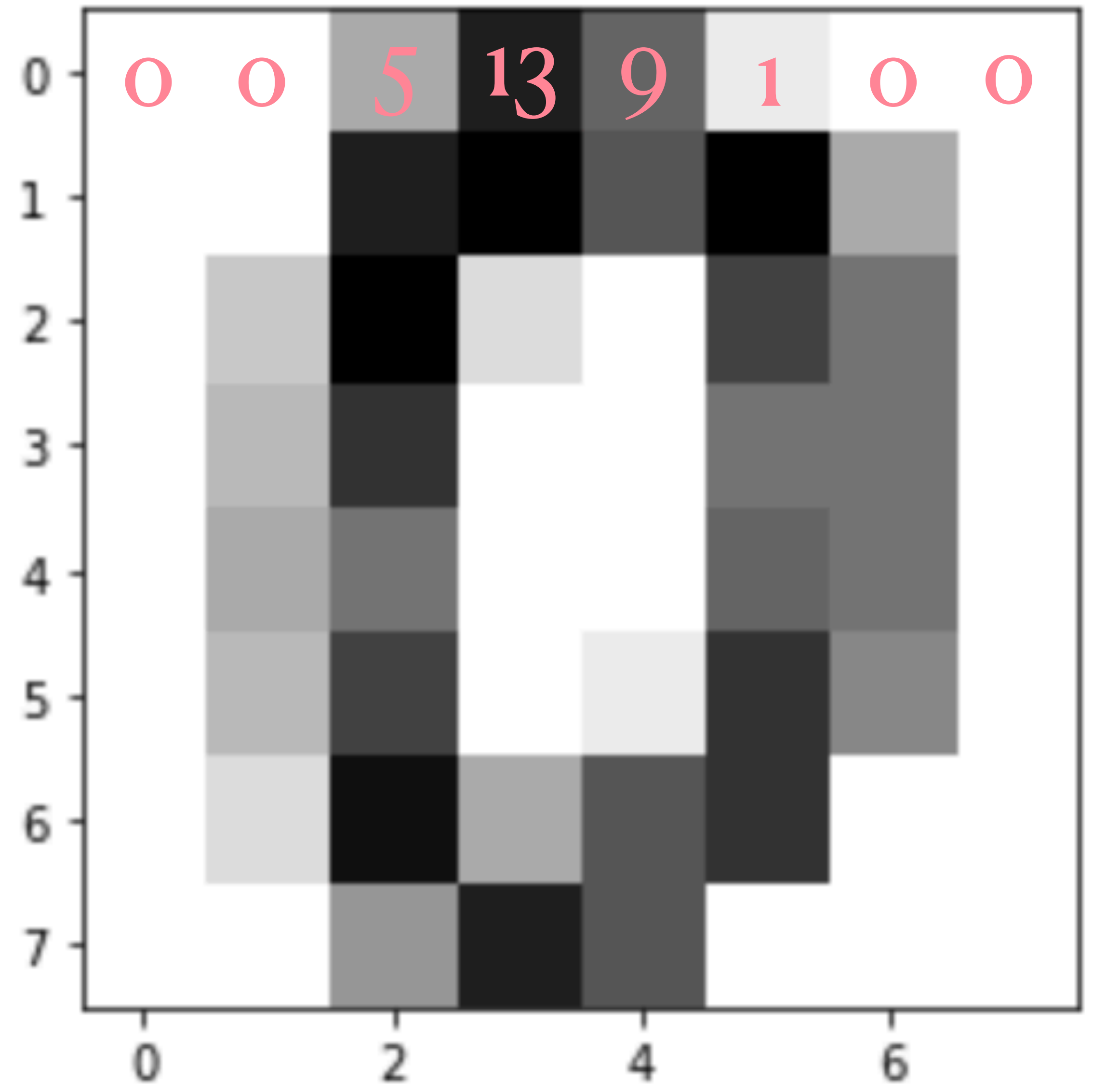
For convenience, we usually do

1. Reshape the image into a 64x1 vector, call it x
2. Call y = 0

See the code in MNIST_basic.ipynb

Model, data set, and loss function

Elements in machine learning




```
digits.images[0]
```

```
array([[ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.],
       [ 0.,  0., 13., 15., 10., 15.,  5.,  0.],
       [ 0.,  3., 15.,  2.,  0., 11.,  8.,  0.],
       [ 0.,  4., 12.,  0.,  0.,  8.,  8.,  0.],
       [ 0.,  5.,  8.,  0.,  0.,  9.,  8.,  0.],
       [ 0.,  4., 11.,  0.,  1., 12.,  7.,  0.],
       [ 0.,  2., 14.,  5., 10., 12.,  0.,  0.],
       [ 0.,  0.,  6., 13., 10.,  0.,  0.,  0.]])
```

< — x

```
# target_names store the label for the digit image

digits.target_names[0]
```

```
0
```

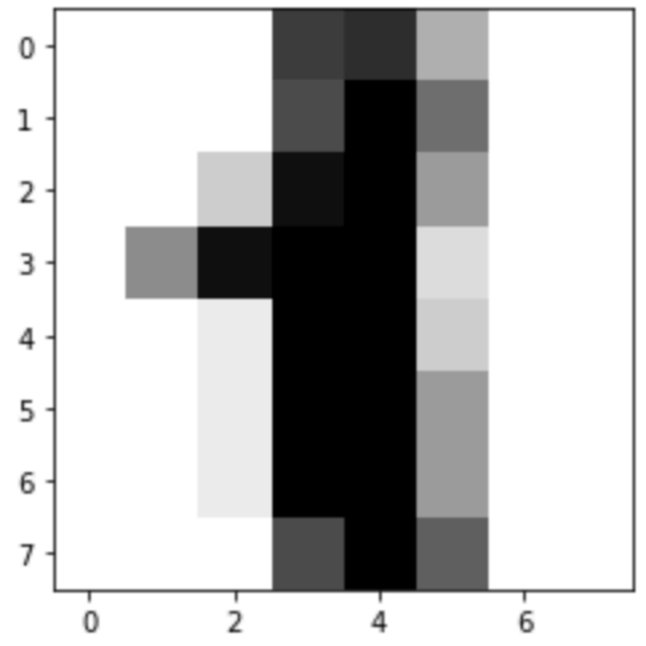
< — y

The whole thing in the left is a datum. The information contained:

1. Image (8x8, 64 numbers)
2. Label '0'

For convenience, we usually do

1. Reshape the image into a 64x1 vector, call it x
2. Call y = 0



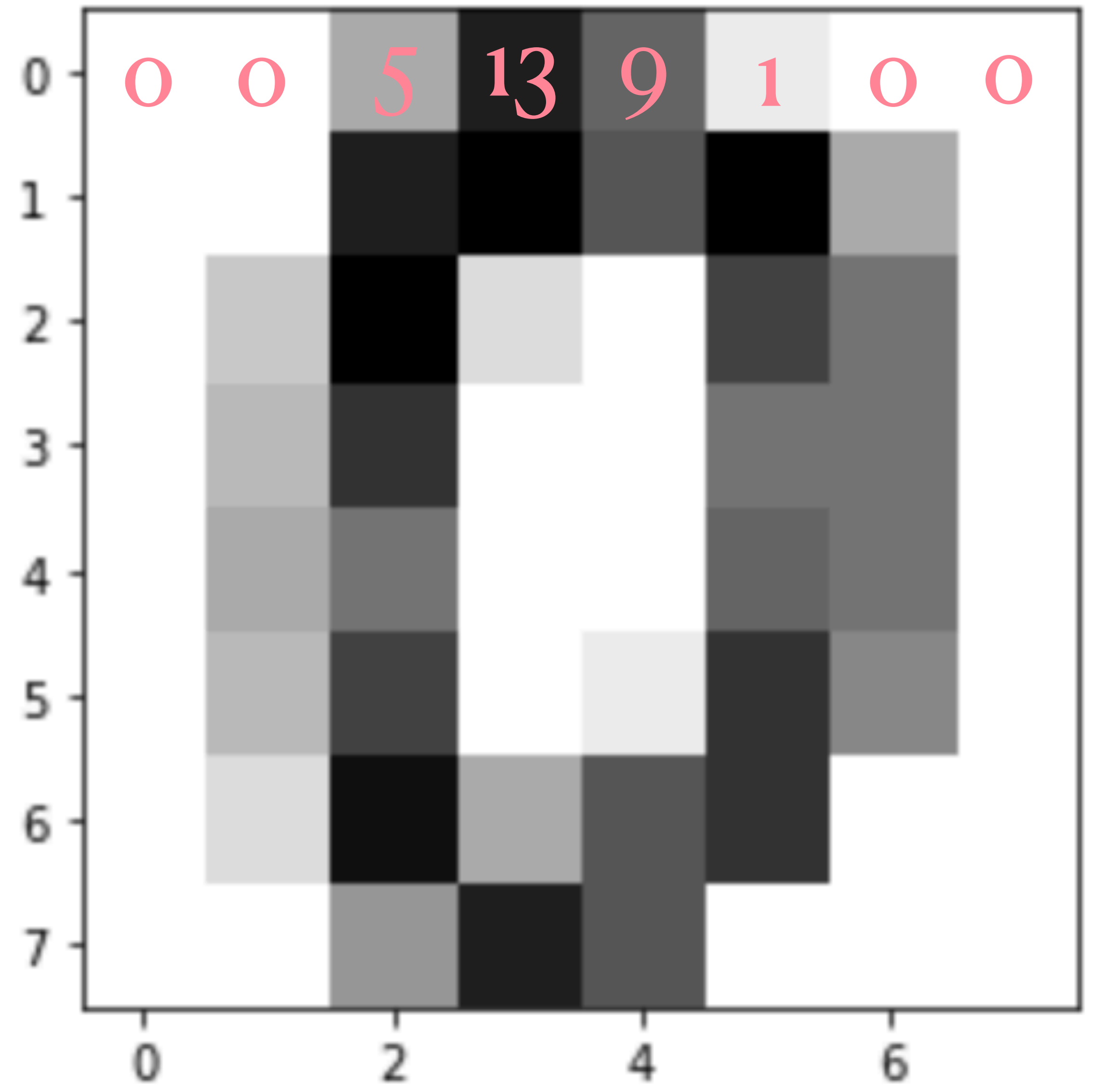
```
digits.images[1]
```

```
array([[ 0.,  0.,  0., 12., 13.,  5.,  0.,  0.],
       [ 0.,  0.,  0., 11., 16.,  9.,  0.,  0.],
       [ 0.,  0.,  3., 15., 16.,  6.,  0.,  0.],
       [ 0.,  7., 15., 16., 16.,  2.,  0.,  0.],
       [ 0.,  0.,  1., 16., 16.,  3.,  0.,  0.],
       [ 0.,  0.,  1., 16., 16.,  6.,  0.,  0.],
       [ 0.,  0.,  1., 16., 16.,  6.,  0.,  0.],
       [ 0.,  0.,  0., 11., 16., 10.,  0.,  0.]])
```

See the code in MNIST_basic.ipynb

Model, data set, and loss function

Elements in machine learning




```
digits.images[0]
```

```
array([[ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.],
       [ 0.,  0., 13., 15., 10., 15.,  5.,  0.],
       [ 0.,  3., 15.,  2.,  0., 11.,  8.,  0.],
       [ 0.,  4., 12.,  0.,  0.,  8.,  8.,  0.],
       [ 0.,  5.,  8.,  0.,  0.,  9.,  8.,  0.],
       [ 0.,  4., 11.,  0.,  1., 12.,  7.,  0.],
       [ 0.,  2., 14.,  5., 10., 12.,  0.,  0.],
       [ 0.,  0.,  6., 13., 10.,  0.,  0.,  0.]])
```

< — x

```
# target_names store the label for the digit image
digits.target_names[0]
```

```
0
```

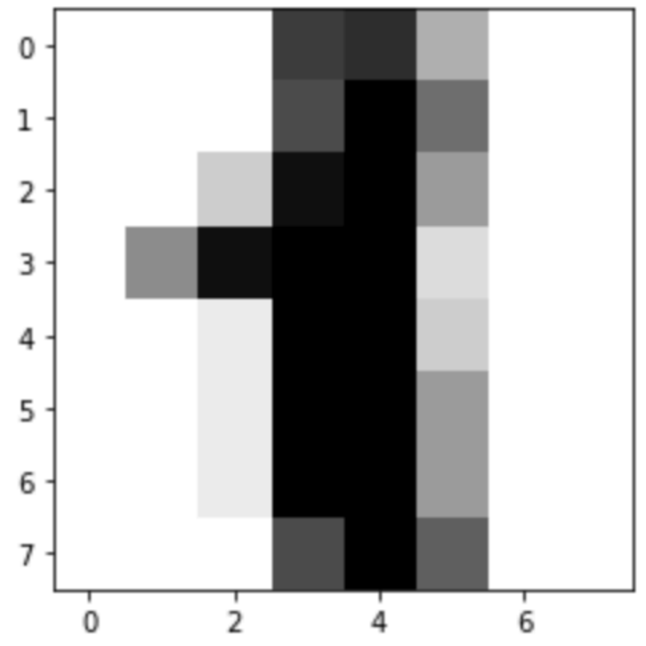
< — y

The whole thing in the left is a datum. The information contained:

1. Image (8x8, 64 numbers)
2. Label '0'

For convenience, we usually do

1. Reshape the image into a 64x1 vector, call it x
2. Call y = 0



```
digits.images[1]
```

```
array([[ 0.,  0.,  0., 12., 13.,  5.,  0.,  0.],
       [ 0.,  0.,  0., 11., 16.,  9.,  0.,  0.],
       [ 0.,  0.,  3., 15., 16.,  6.,  0.,  0.],
       [ 0.,  7., 15., 16., 16.,  2.,  0.,  0.],
       [ 0.,  0.,  1., 16., 16.,  3.,  0.,  0.],
       [ 0.,  0.,  1., 16., 16.,  6.,  0.,  0.],
       [ 0.,  0.,  1., 16., 16.,  6.,  0.,  0.],
       [ 0.,  0.,  0., 11., 16., 10.,  0.,  0.]])
```

```
digits.target_names[1]
```

```
1
```

See the code in MNIST_basic.ipynb

Mathematical notation for data set

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}. \qquad (1.1)$$

Mathematical notation for data set

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}. \qquad (1.1)$$

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

Mathematical notation for data set

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}. \tag{1.1}$$

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

Mathematical notation for data set

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}. \qquad (1.1)$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

Mathematical notation for data set

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}. \qquad (1.1)$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix}$$

Interpretation: there are n data, each datum contains p features (pixels in MNIST)

Mathematical notation for data set

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}. \qquad (1.1)$$

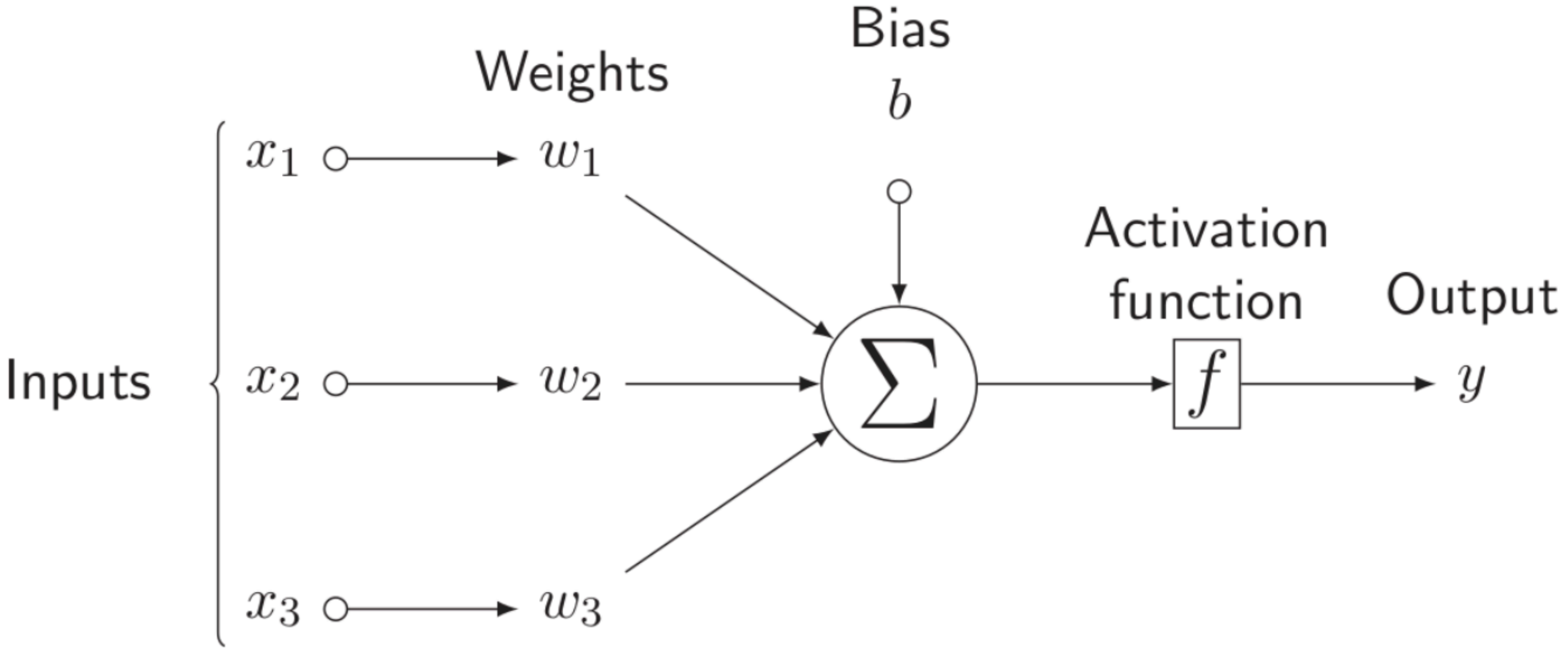
$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ \underline{x_{21}} & \underline{x_{22}} & \underline{\dots} & \underline{x_{2p}} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

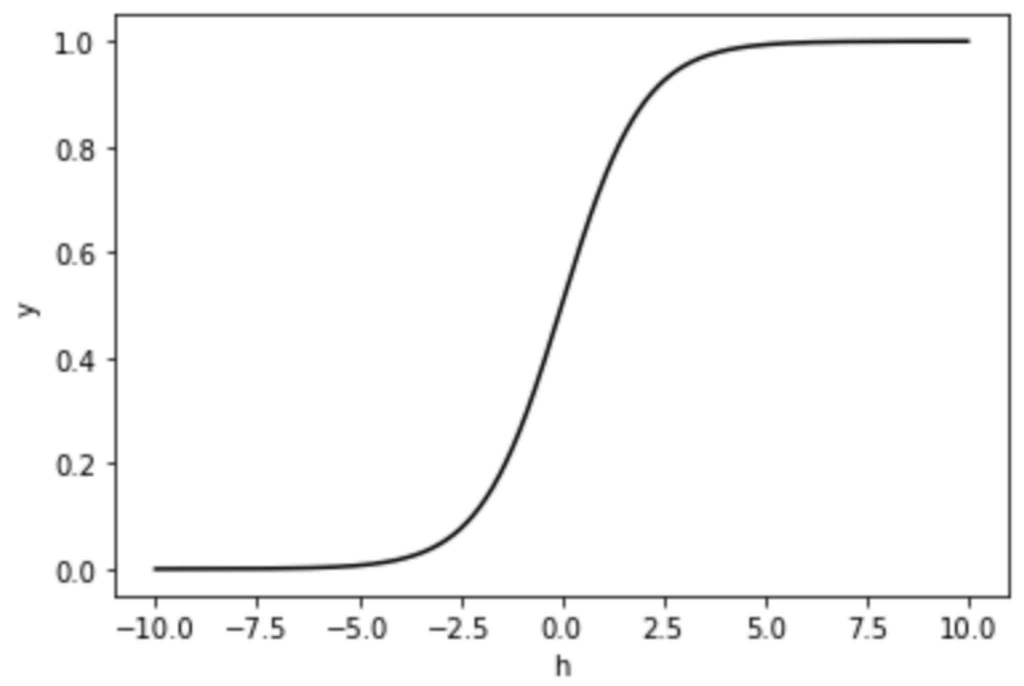
$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Interpretation: there are n data, each datum contains p features (pixels in MNIST)

$\boxed{\text{The collection \{X, y\} is a data set.}}$

Model: a mathematical form for $y = f(\mathbf{x} \mid w, b)$



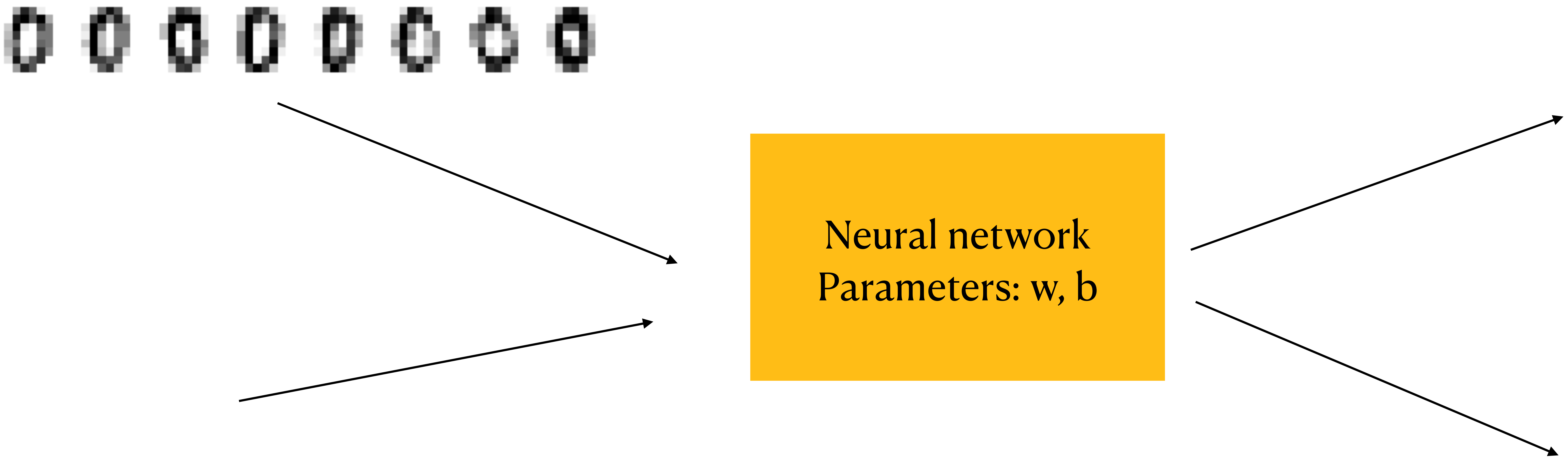


https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network

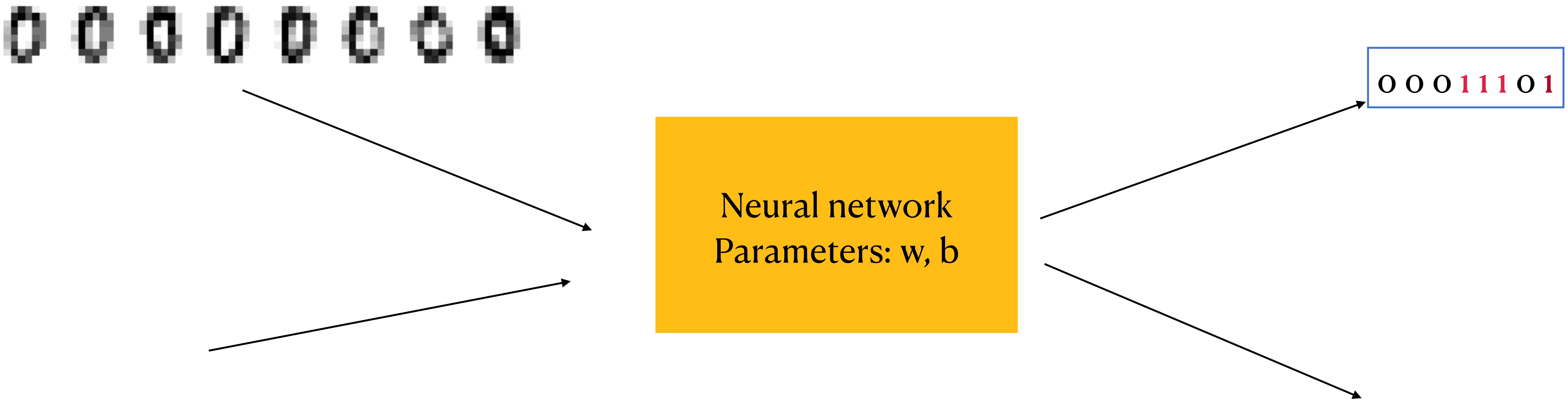
$$h(\mathbf{x}) = b + \sum_i w_i x_i$$

$$f(h) = \frac{1}{1 + e^{-h}}$$

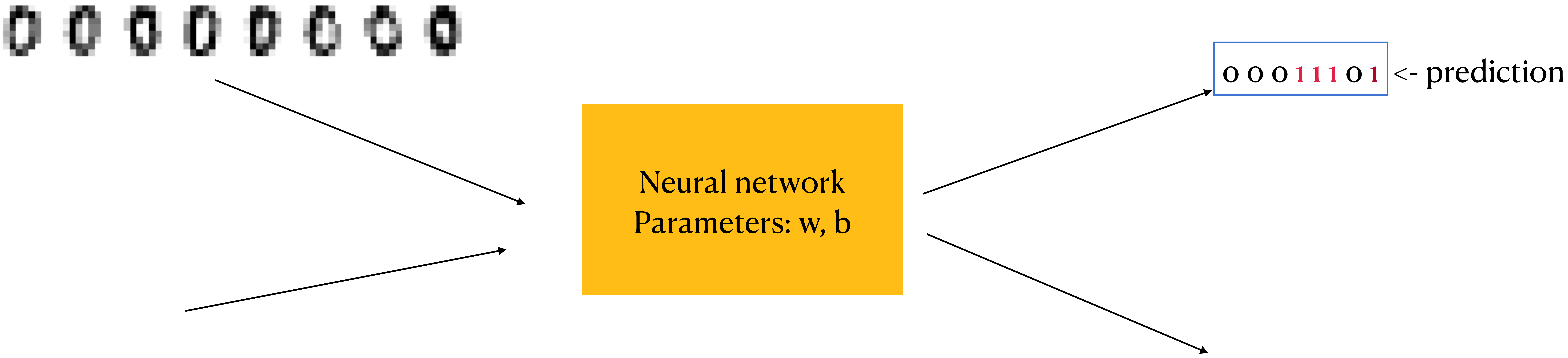
(Qualitative) lose function



(Qualitative) lose function



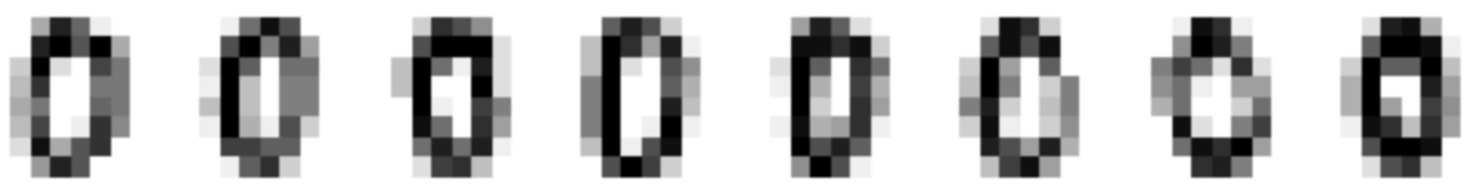
(Qualitative) lose function



(Qualitative) lose function

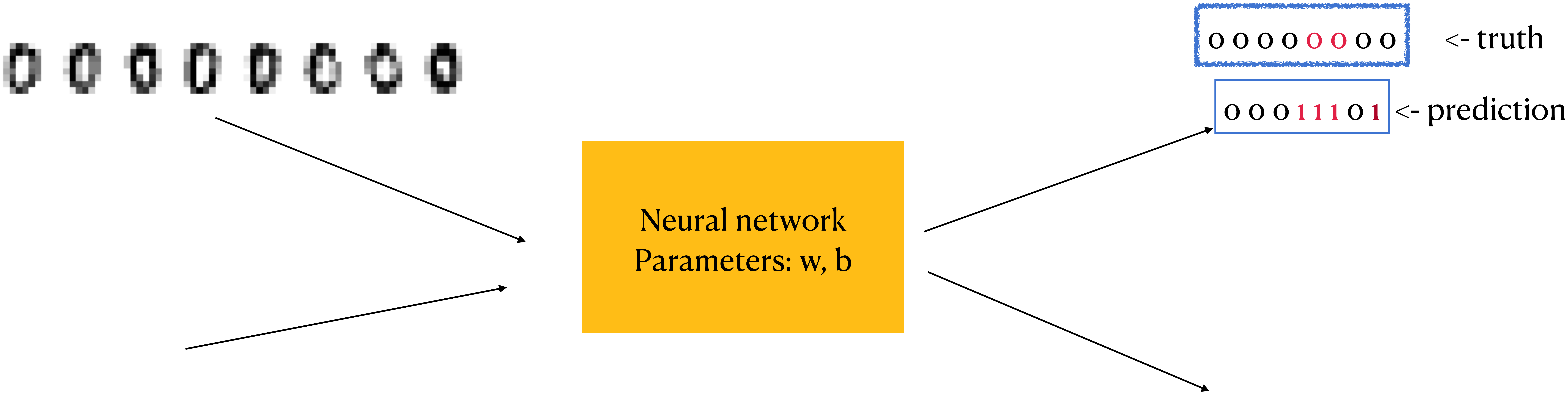
<- truth

0 0 0 1 1 1 0 1 <- prediction

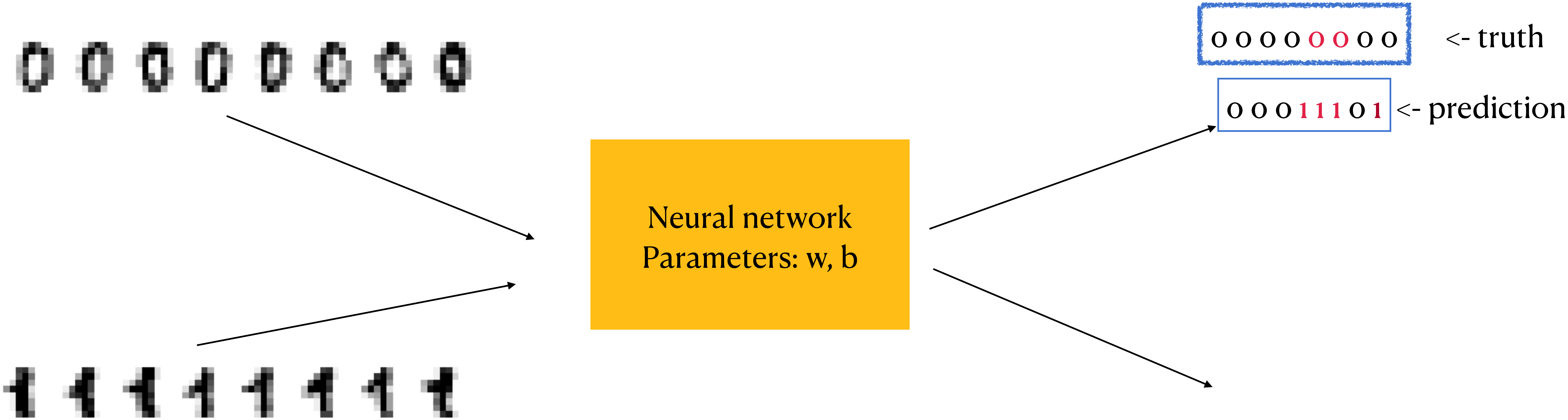


Neural network
Parameters: w, b

(Qualitative) lose function



(Qualitative) lose function

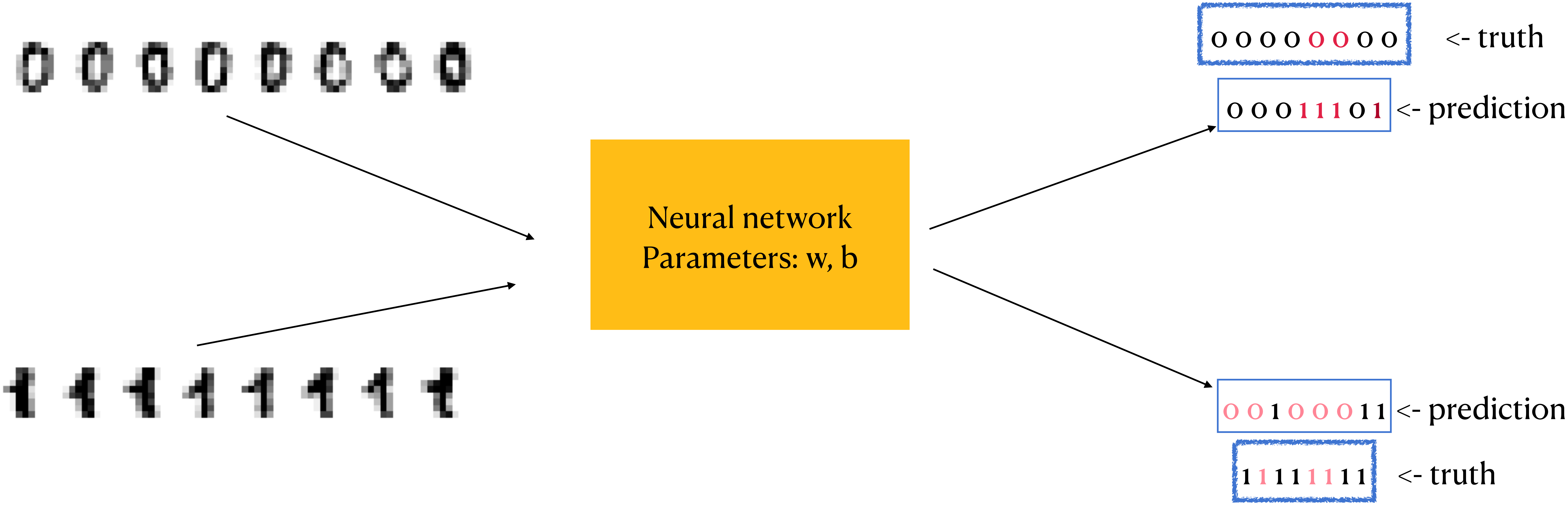


(Qualitative) lose function

(Qualitative) lose function

(Qualitative) lose function

(Qualitative) lose function

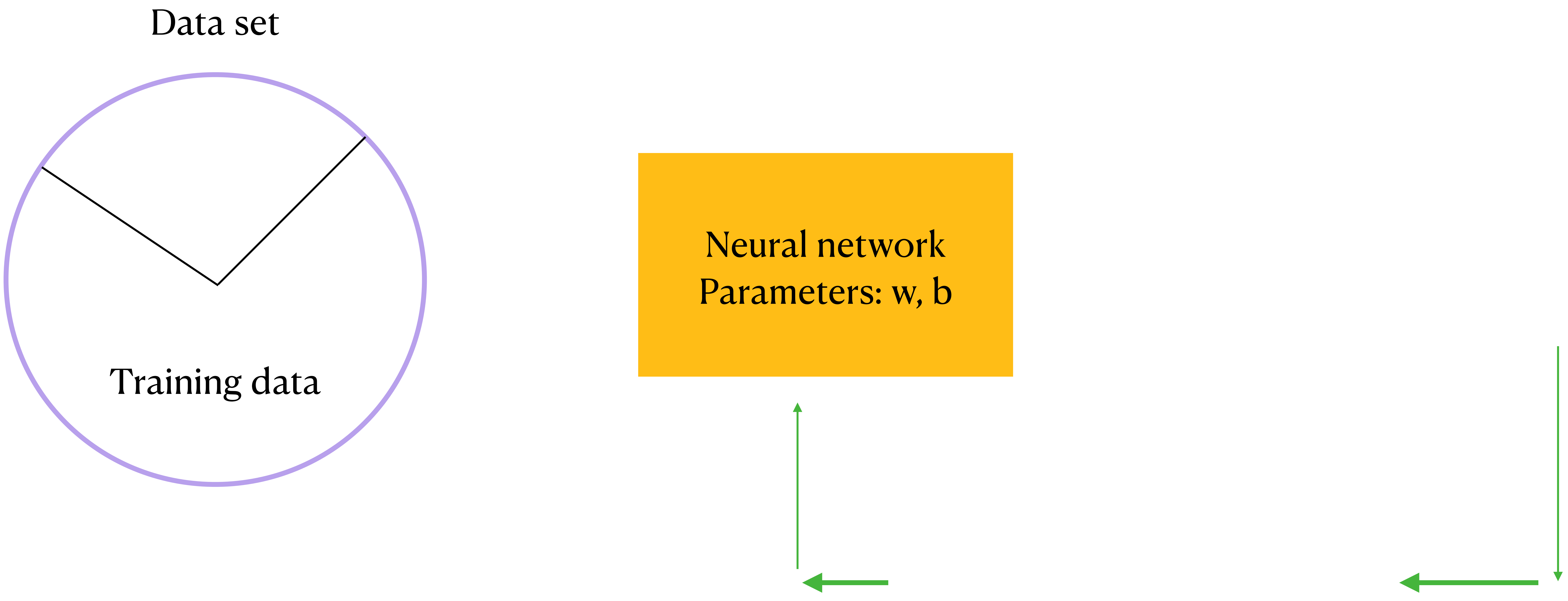


(Qualitative) lose function

Total number of mistakes: 4 + 5 = 9 out of 16 samples

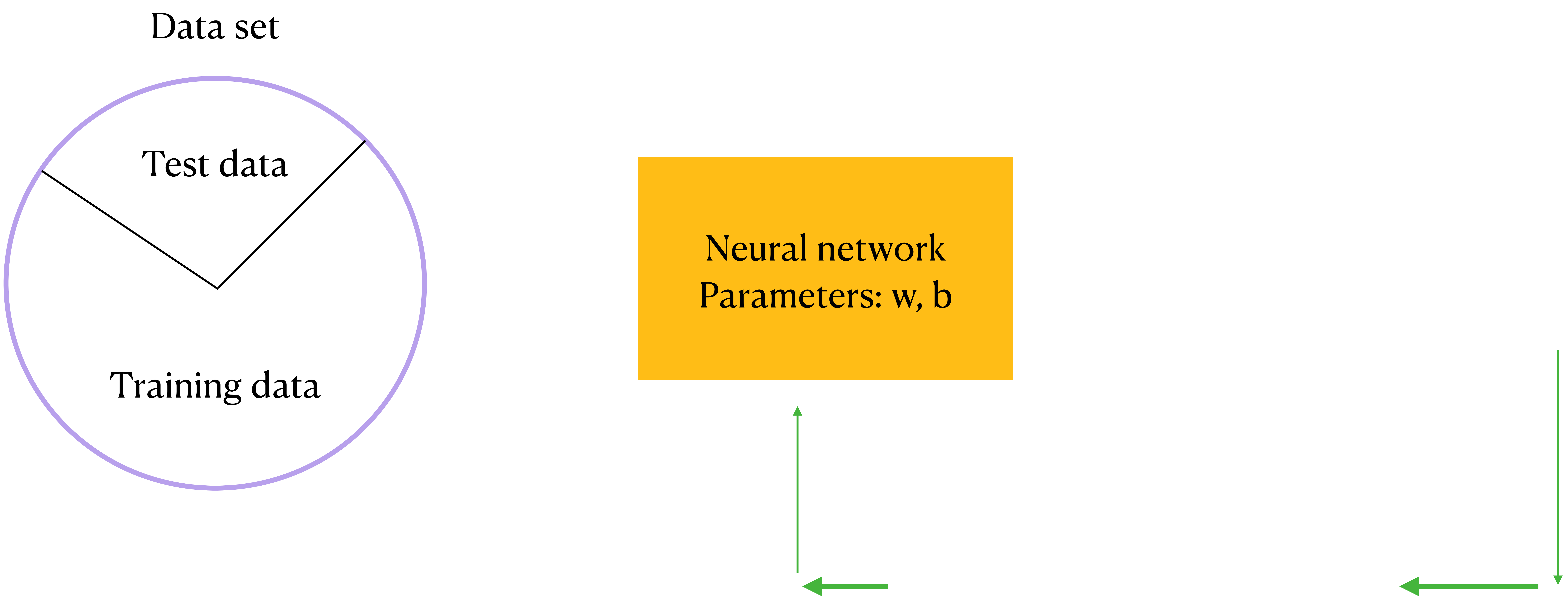
So what is machine learning?

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



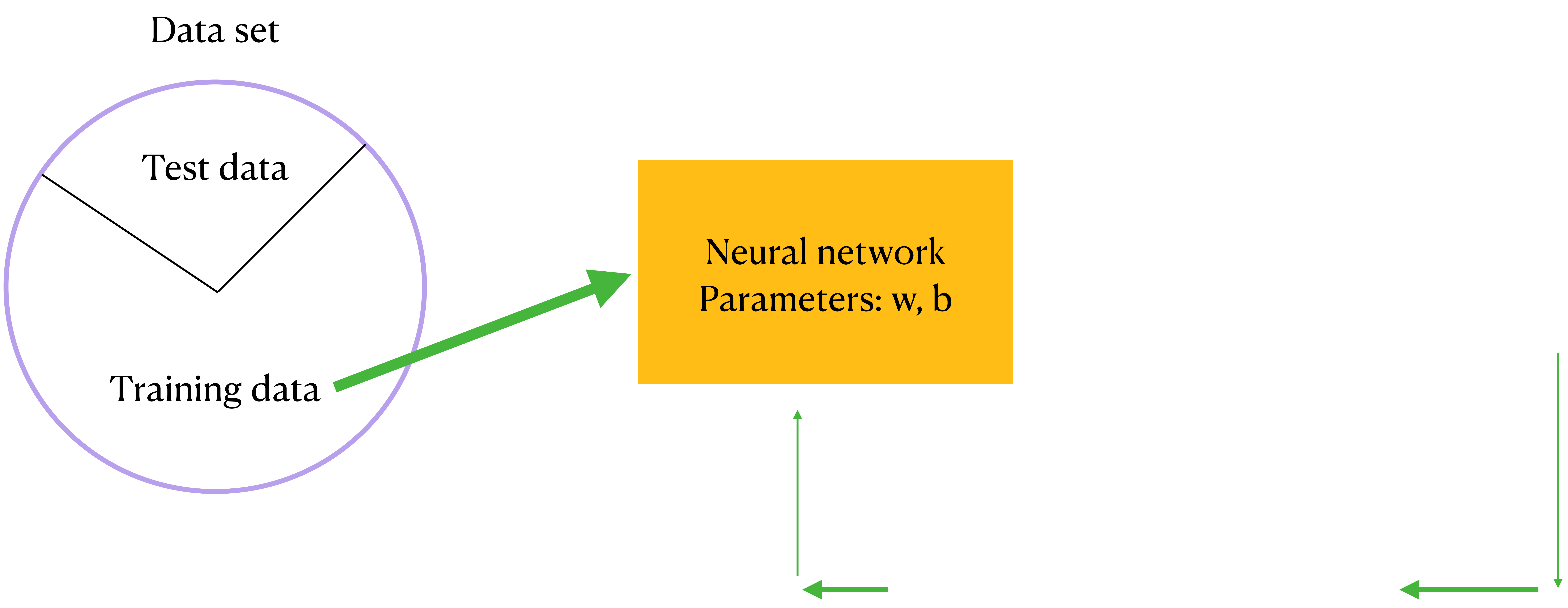
So what is machine learning?

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



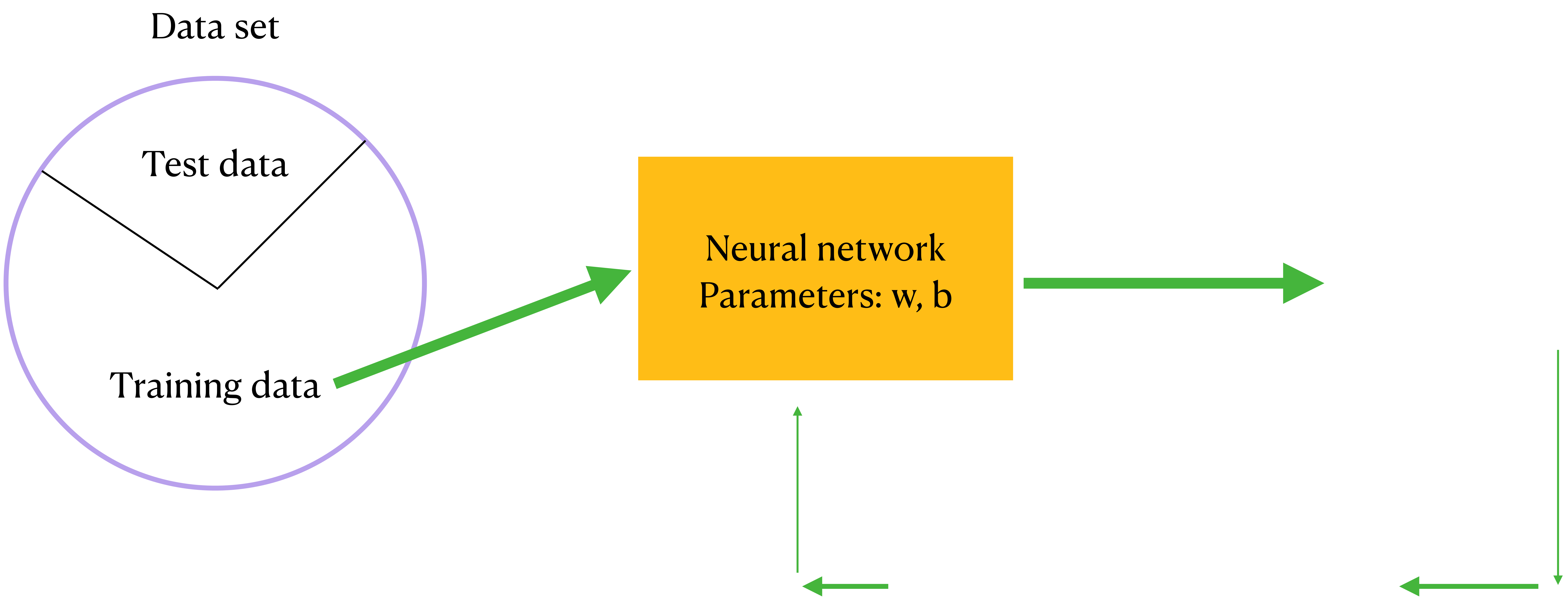
So what is machine learning?

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



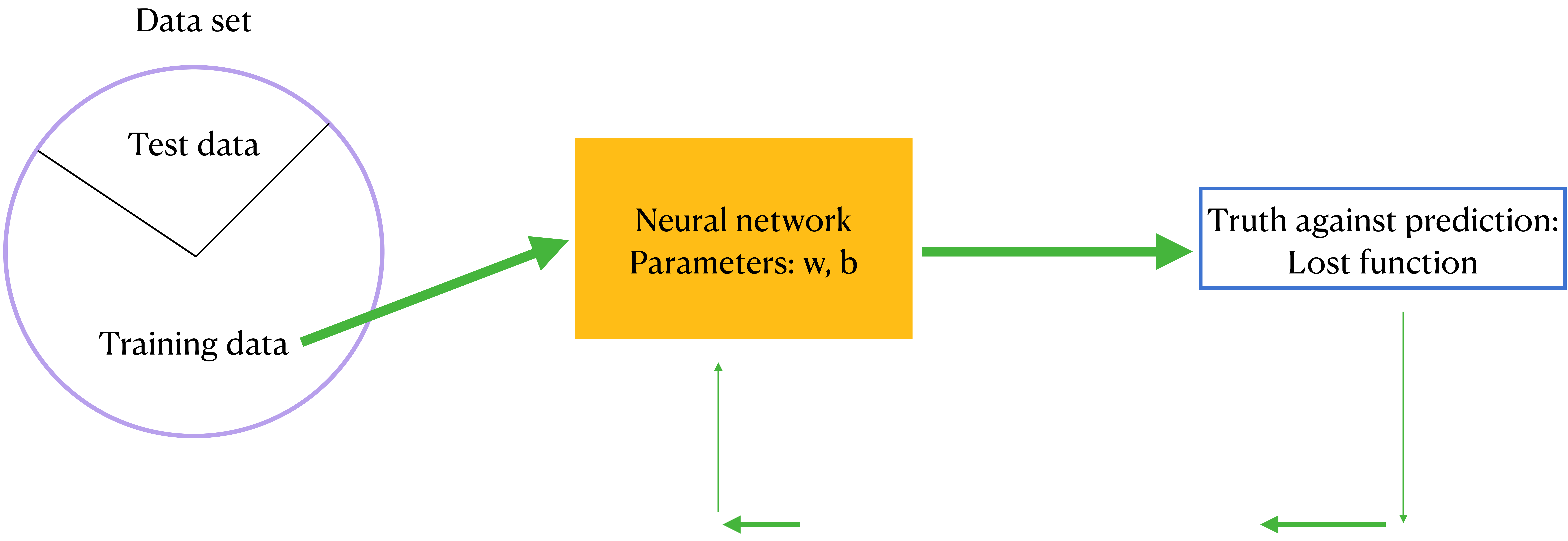
So what is machine learning?

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



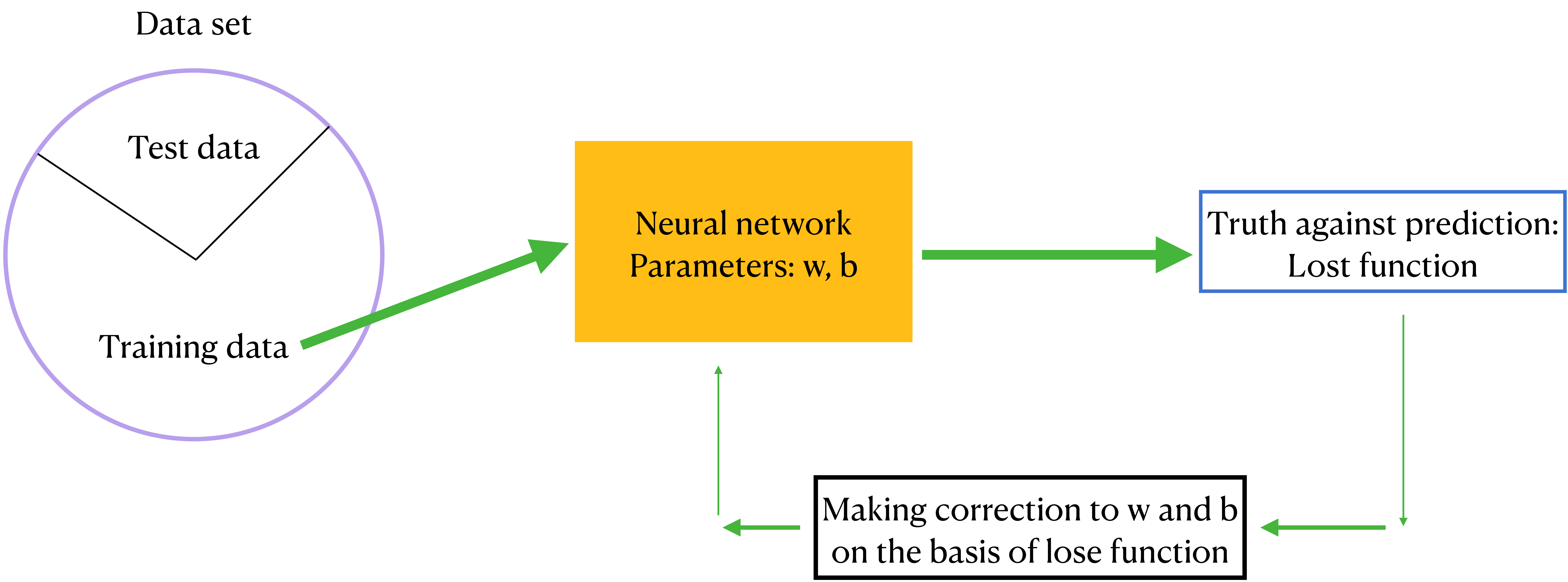
So what is machine learning?

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



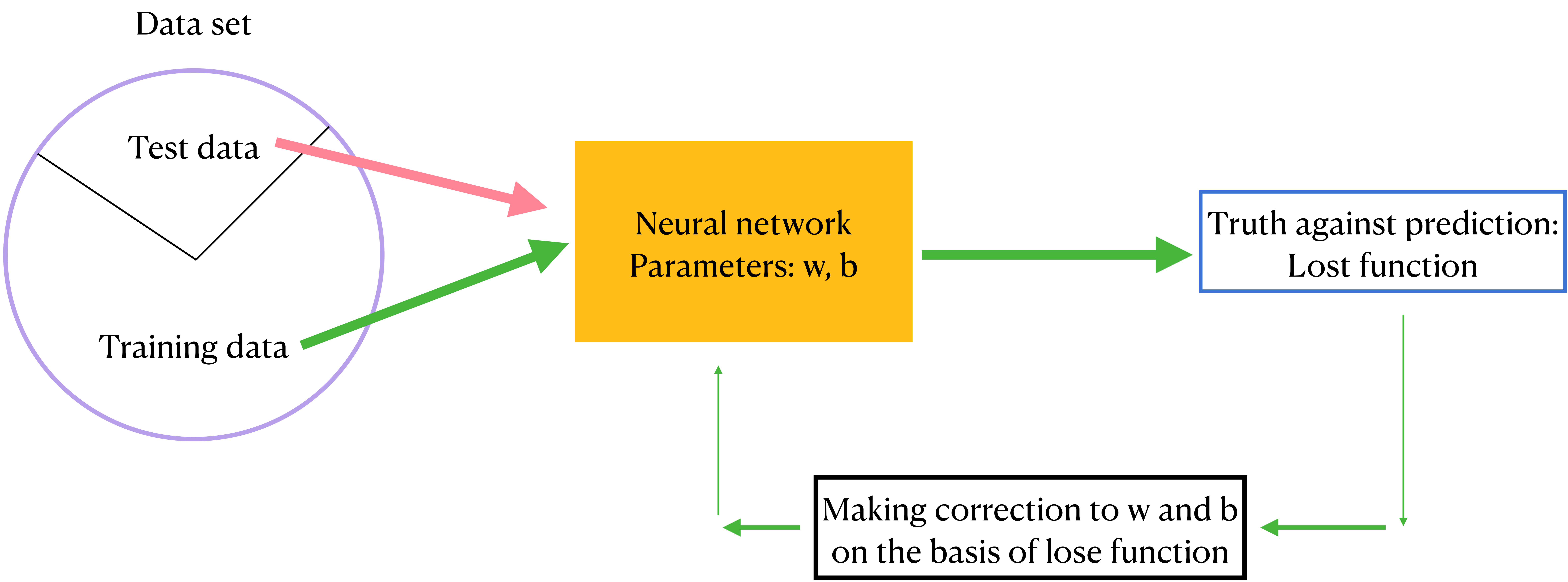
So what is machine learning?

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



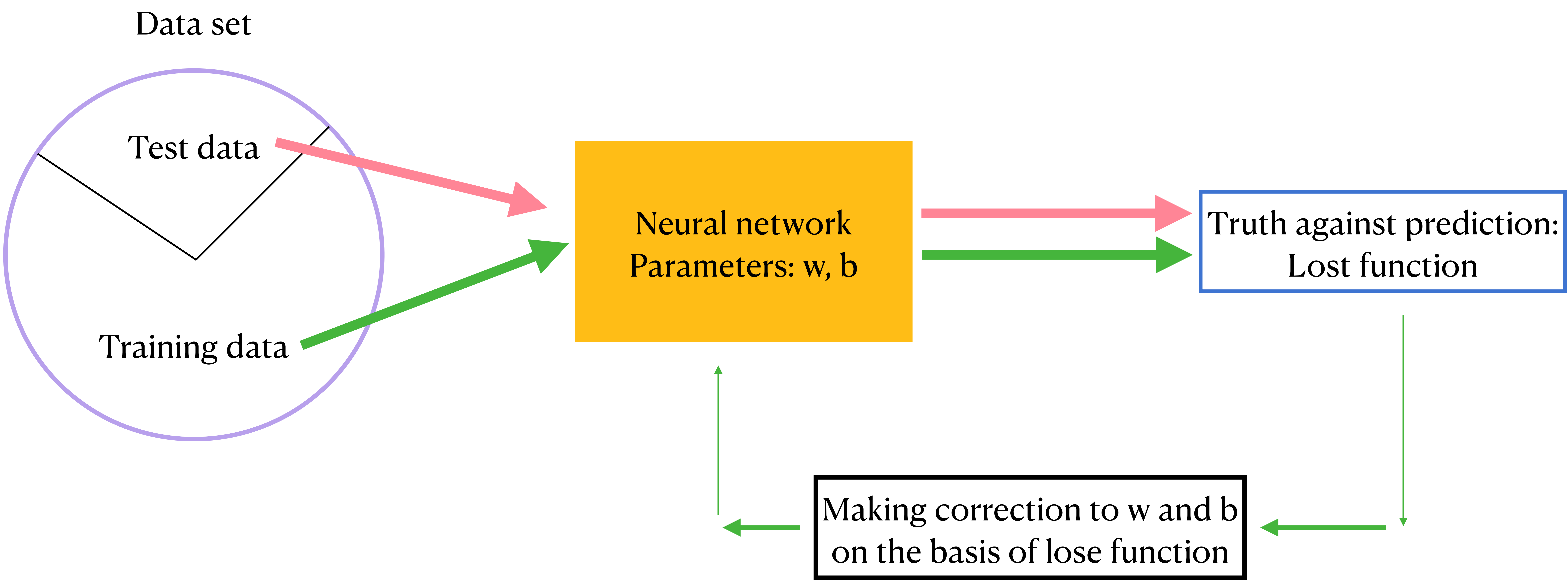
So what is machine learning?

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



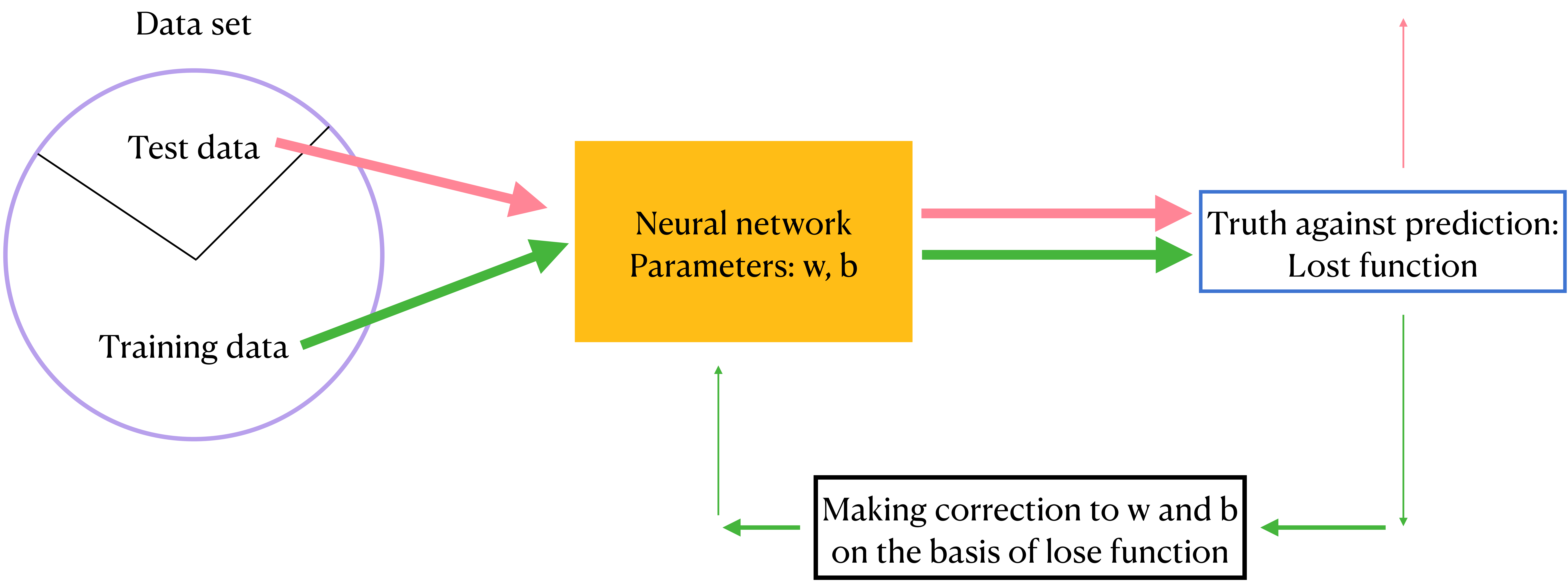
So what is machine learning?

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



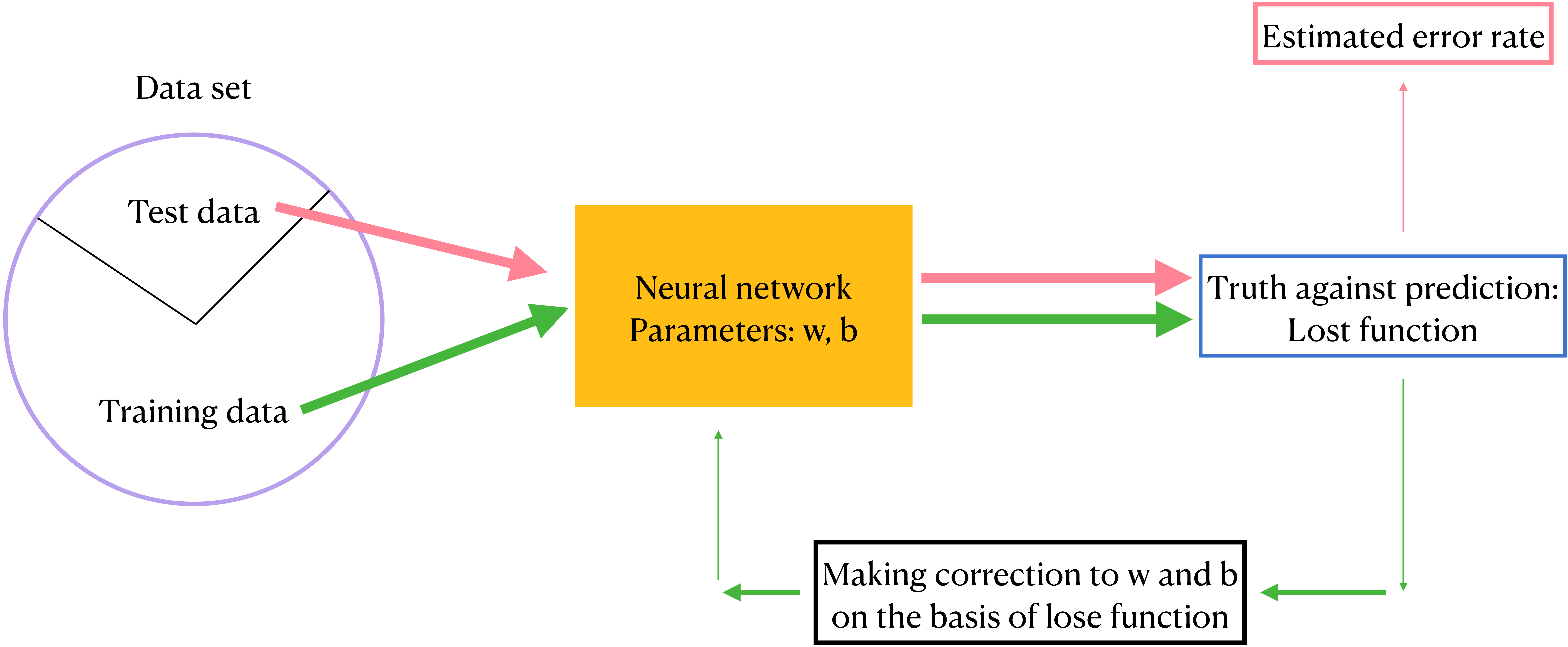
So what is machine learning?

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



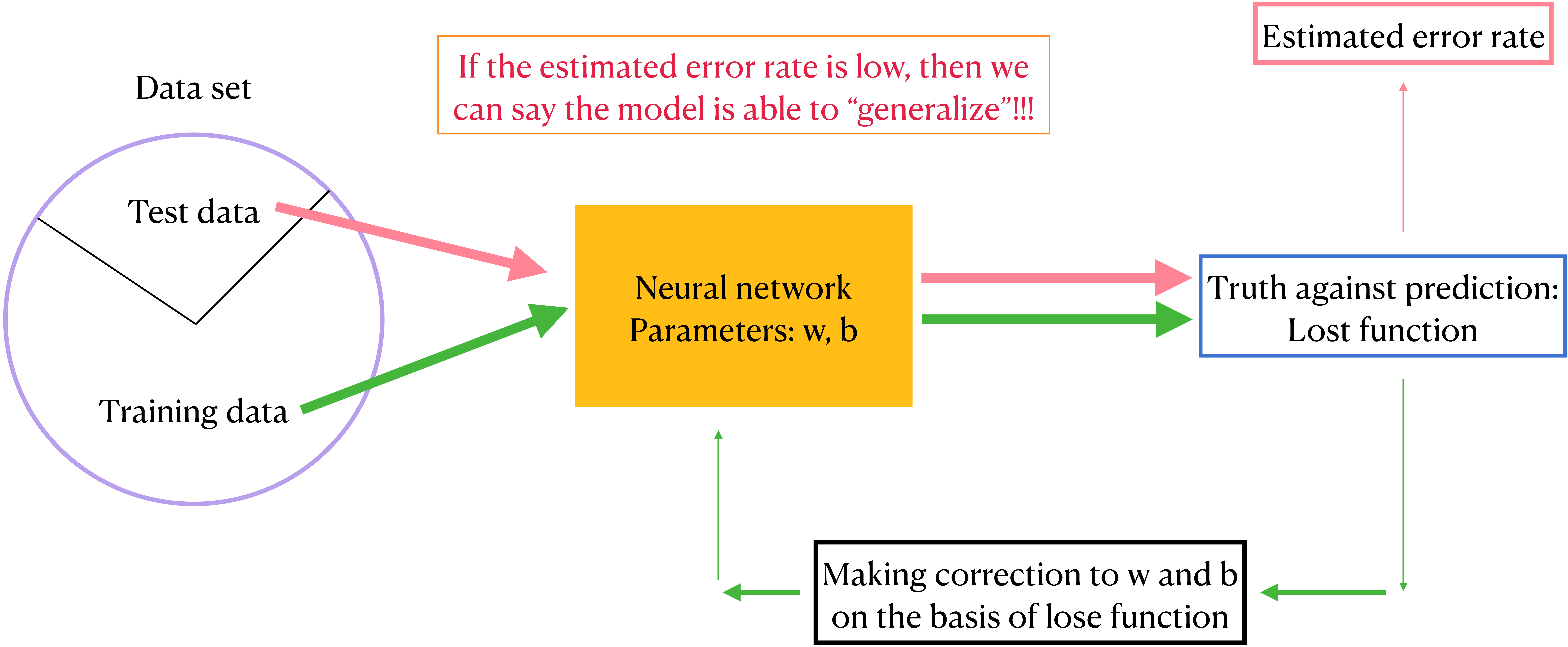
So what is machine learning?

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.



Important notes for supervised learning

Must review this slide before midterm

Important notes for supervised learning

Must review this slide before midterm

Model is a mathematic function $\hat{y} = f(x \mid w, b)$ which takes input x of a datum and output a prediction $\hat{y}$

Important notes for supervised learning

Must review this slide before midterm

Model is a mathematic function $\hat{y} = f(x \mid w, b)$ which takes input x of a datum and output a prediction $\hat{y}$

Lose function measures the deviation between the truth y and the prediction $\hat{y}$

Important notes for supervised learning

Must review this slide before midterm

Model is a mathematic function $\hat{y} = f(x \,|\, w, b)$ which takes input x of a datum and output a prediction $\hat{y}$

Lose function measures the deviation between the truth y and the prediction $\hat{y}$

Training error means the average lose function values associated with the training data, i.e. measure of average difference between y_{train} and $\hat{y}_{train} = f(x_{train} \,|\, w, b)$

Important notes for supervised learning

Must review this slide before midterm

Model is a mathematic function $\hat{y} = f(x \mid w, b)$ which takes input x of a datum and output a prediction $\hat{y}$

Lose function measures the deviation between the truth y and the prediction $\hat{y}$

Training error means the average lose function values associated with the training data, i.e. measure of average difference between y_{train} and $\hat{y}_{train} = f(x_{train} \mid w, b)$

Machine learning use some algorithm based on training error to correct the model parameters w and b.

Important notes for supervised learning

Must review this slide before midterm

Model is a mathematic function $\hat{y} = f(x \mid w, b)$ which takes input x of a datum and output a prediction $\hat{y}$

Lose function measures the deviation between the truth y and the prediction $\hat{y}$

Training error means the average lose function values associated with the training data, i.e. measure of average difference between y_{train} and $\hat{y}_{train} = f(x_{train} \mid w, b)$

Machine learning use some algorithm based on training error to correct the model parameters w and b.

So during the training, the training error changes with time, and should be on the decreasing tracking.

Important notes for supervised learning

Must review this slide before midterm

Model is a mathematic function $\hat{y} = f(x \mid w, b)$ which takes input x of a datum and output a prediction $\hat{y}$

Lose function measures the deviation between the truth y and the prediction $\hat{y}$

Training error means the average lose function values associated with the training data, i.e. measure of average difference between y_{train} and $\hat{y}_{train} = f(x_{train} \mid w, b)$

Machine learning use some algorithm based on training error to correct the model parameters w and b.

So during the training, the training error changes with time, and should be on the decreasing tracking.

So during the training, the model parameters w and b are varying too.

Important notes for supervised learning

Must review this slide before midterm

Model is a mathematic function $\hat{y} = f(x \mid w, b)$ which takes input x of a datum and output a prediction $\hat{y}$

Lose function measures the deviation between the truth y and the prediction $\hat{y}$

Training error means the average lose function values associated with the training data, i.e. measure of average difference between y_{train} and $\hat{y}_{train} = f(x_{train} \mid w, b)$

Machine learning use some algorithm based on training error to correct the model parameters w and b.

So during the training, the training error changes with time, and should be on the decreasing tracking.

So during the training, the model parameters w and b are varying too.

Once the optimal $\bar{w}$ and $\bar{b}$ are reached, the training is complete.

Important notes for supervised learning

Must review this slide before midterm

Model is a mathematic function $\hat{y} = f(x \,|\, w, b)$ which takes input x of a datum and output a prediction $\hat{y}$

Lose function measures the deviation between the truth y and the prediction $\hat{y}$

Training error means the average lose function values associated with the training data, i.e. measure of average difference between y_{train} and $\hat{y}_{train} = f(x_{train} \,|\, w, b)$

Machine learning use some algorithm based on training error to correct the model parameters w and b.

So during the training, the training error changes with time, and should be on the decreasing tracking.

So during the training, the model parameters w and b are varying too.

Once the optimal $\bar{w}$ and $\bar{b}$ are reached, the training is complete.

Now we feed the test data into the trained model, we calculate the test error be comparing y_{test} against $\hat{y}_{test} = f(x_{test} \,|\, \bar{w}, \bar{b})$

Important notes for supervised learning

Must review this slide before midterm

Model is a mathematic function $\hat{y} = f(x \mid w, b)$ which takes input x of a datum and output a prediction $\hat{y}$

Lose function measures the deviation between the truth y and the prediction $\hat{y}$

Training error means the average lose function values associated with the training data, i.e. measure of average difference between y_{train} and $\hat{y}_{train} = f(x_{train} \mid w, b)$

Machine learning use some algorithm based on training error to correct the model parameters w and b.

So during the training, the training error changes with time, and should be on the decreasing tracking.

So during the training, the model parameters w and b are varying too.

Once the optimal $\bar{w}$ and $\bar{b}$ are reached, the training is complete.

Now we feed the test data into the trained model, we calculate the test error be comparing y_{test} against $\hat{y}_{test} = f(x_{test} \mid \bar{w}, \bar{b})$

What you should deliver to the customer: model $f(x \mid w, b)$, optimal parameter $\bar{w}$ and $\bar{b}$, and the test error.

Important notes for supervised learning

Must review this slide before midterm

Model is a mathematic function $\hat{y} = f(x \mid w, b)$ which takes input x of a datum and output a prediction $\hat{y}$

Lose function measures the deviation between the truth y and the prediction $\hat{y}$

Training error means the average lose function values associated with the training data, i.e. measure of average difference between y_{train} and $\hat{y}_{train} = f(x_{train} \mid w, b)$

Machine learning use some algorithm based on training error to correct the model parameters w and b.

So during the training, the training error changes with time, and should be on the decreasing tracking.

So during the training, the model parameters w and b are varying too.

Once the optimal $\bar{w}$ and $\bar{b}$ are reached, the training is complete.

Now we feed the test data into the trained model, we calculate the test error be comparing y_{test} against $\hat{y}_{test} = f(x_{test} \mid \bar{w}, \bar{b})$

What you should deliver to the customer: model $f(x \mid w, b)$, optimal parameter $\bar{w}$ and $\bar{b}$, and the test error.

Supervised learning versus unsupervised learning

It is subtle, but an easy answer will be …

Supervised learning:
e.g. Image classification
Data set: {X, y}

Supervised learning versus unsupervised learning

It is subtle, but an easy answer will be ...

Supervised learning:
e.g. Image classification
Data set: {X, y}

unsupervised learning:
e.g. Movie recommender, text auto-completion
Data set: {X}

Supervised learning versus unsupervised learning

It is subtle, but an easy answer will be ...

Supervised learning:
e.g. Image classification
Data set: {X, y}

unsupervised learning:
e.g. Movie recommender, text auto-completion
Data set: {X}

Learning is driven by loss function.
Metaphor: physical penalty

Supervised learning versus unsupervised learning

It is subtle, but an easy answer will be ...

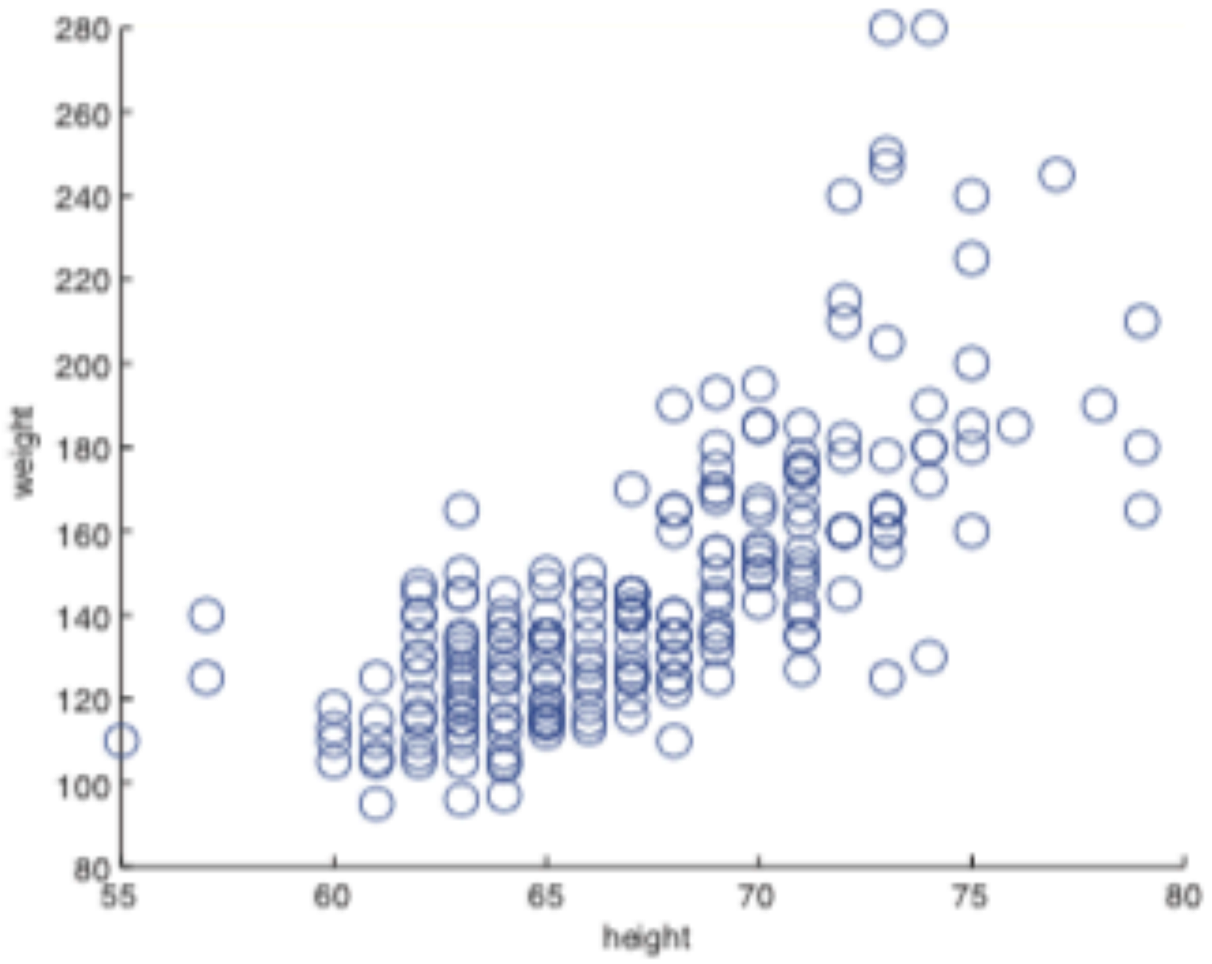
Supervised learning:
e.g. Image classification
Data set: {X, y}

unsupervised learning:
e.g. Movie recommender, text auto-completion
Data set: {X}

Learning is driven by loss function.
Metaphor: physical penalty

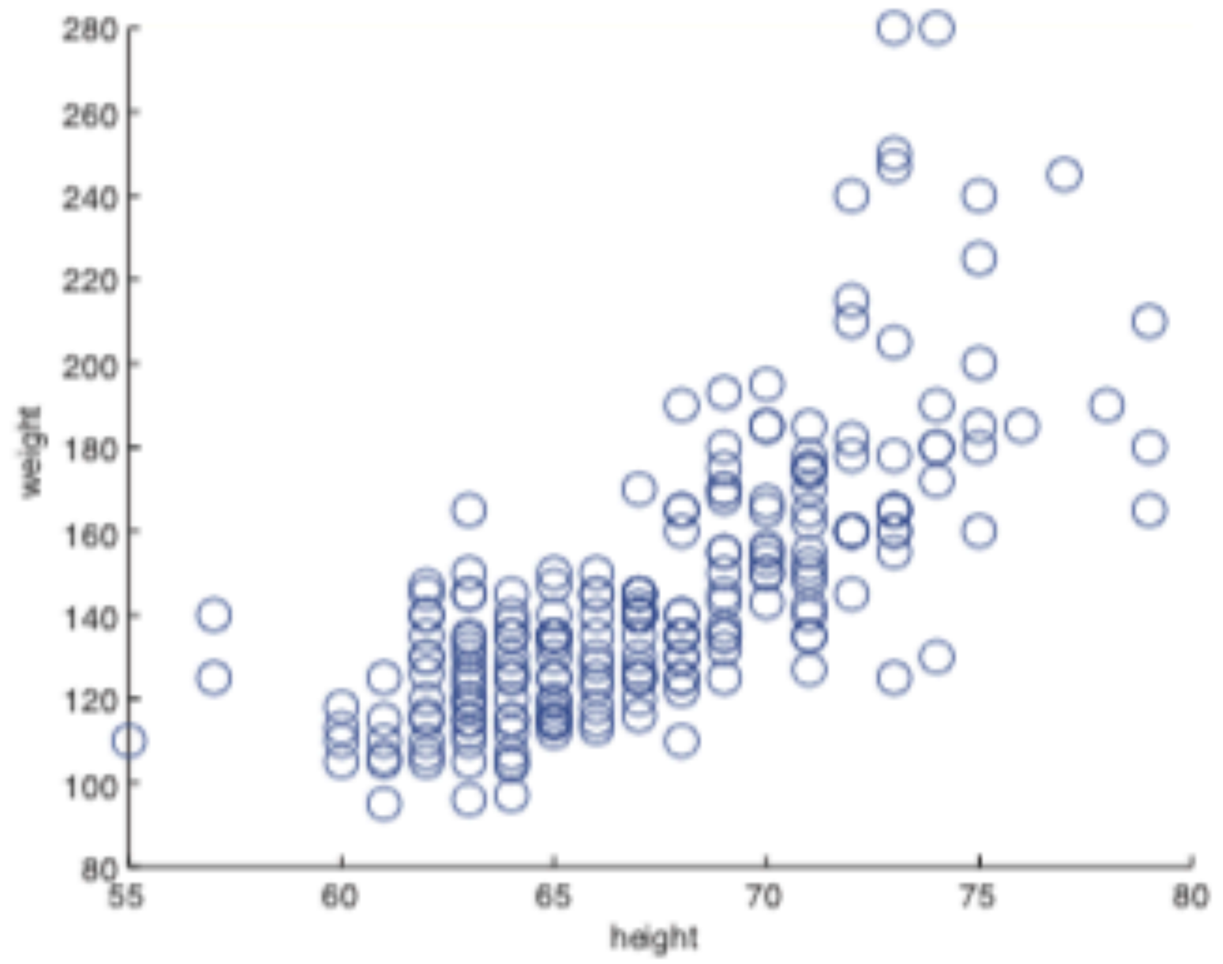
Learning is driven by likelihood.
Metaphor: education with love

Task: given height-weight of a group of students, to separate data into two subgroups

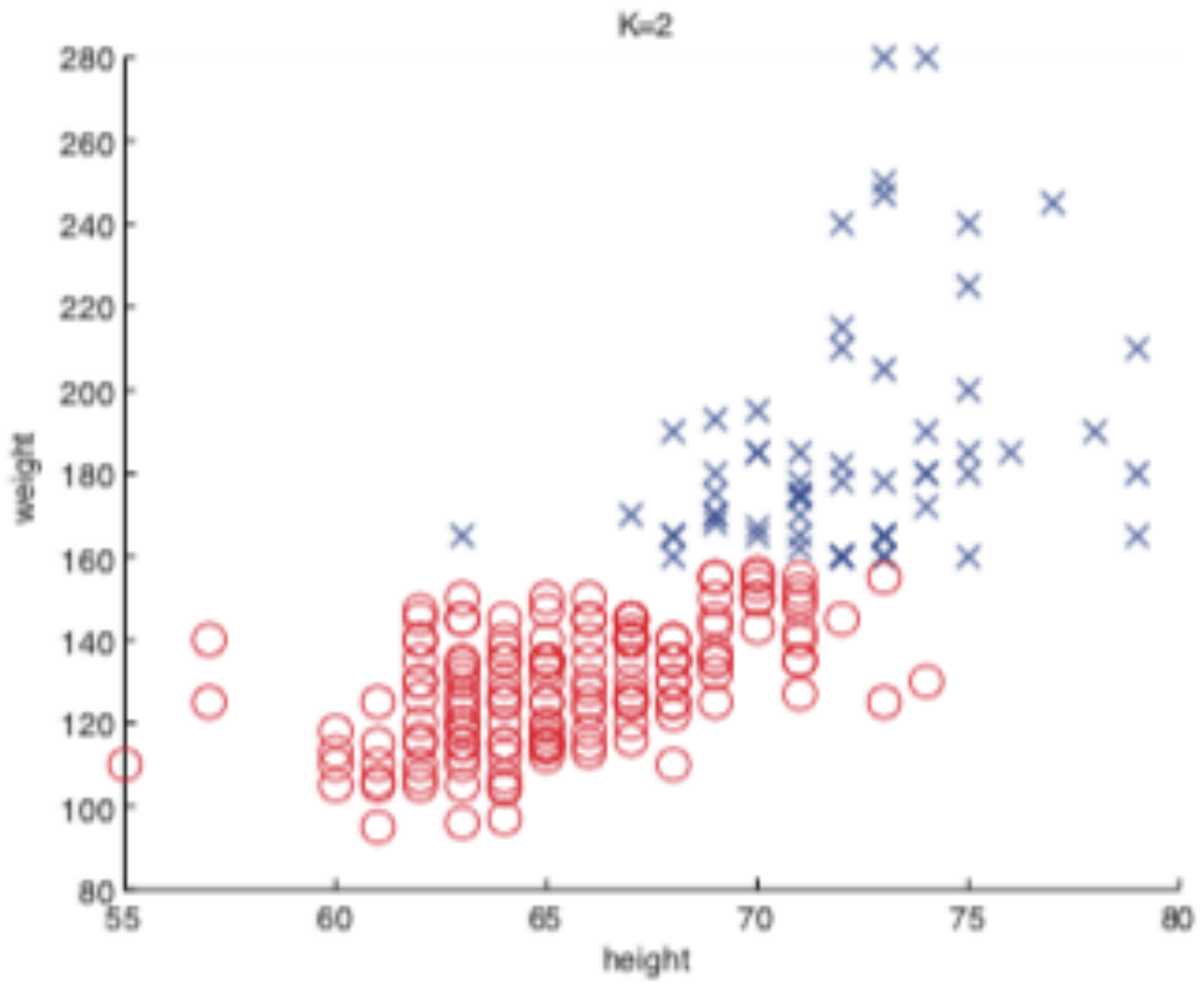


(a)

Task: given height-weight of a group of students, to separate data into two subgroups



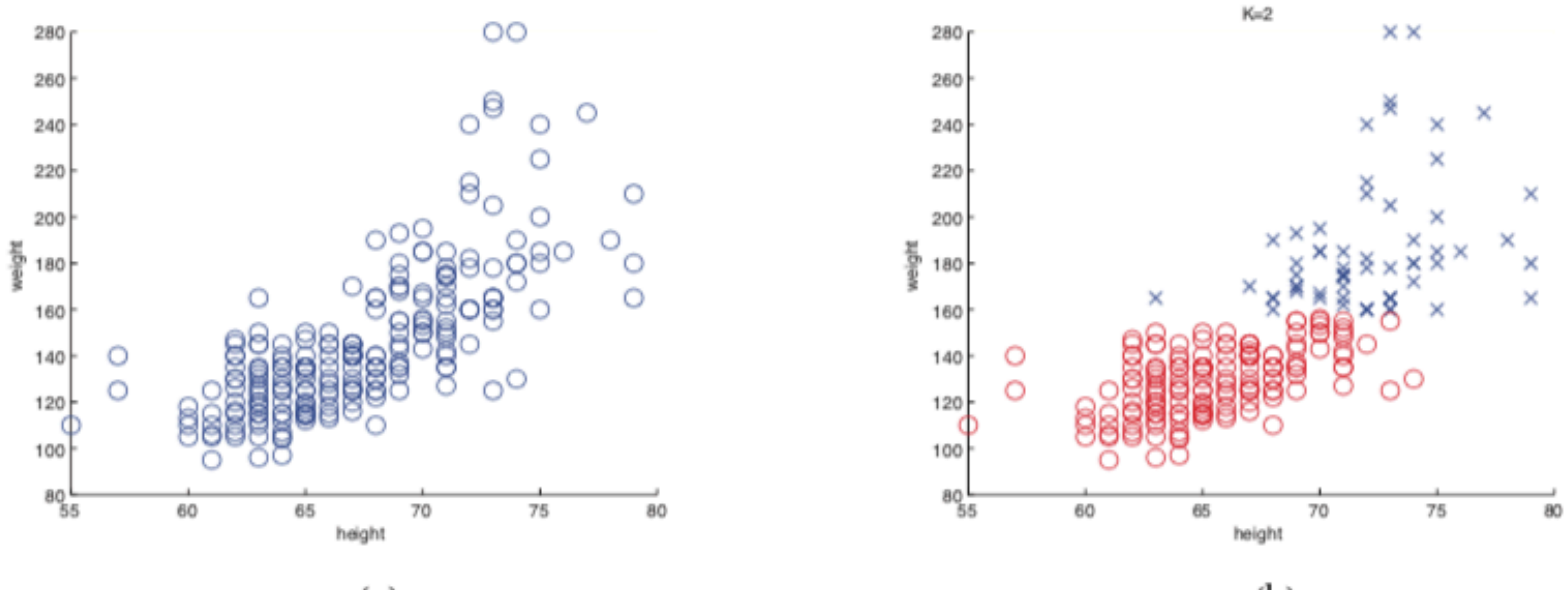
(a)



(b)

clustering

Task: given height-weight of a group of students, to separate data into two subgroups



(a) (b)

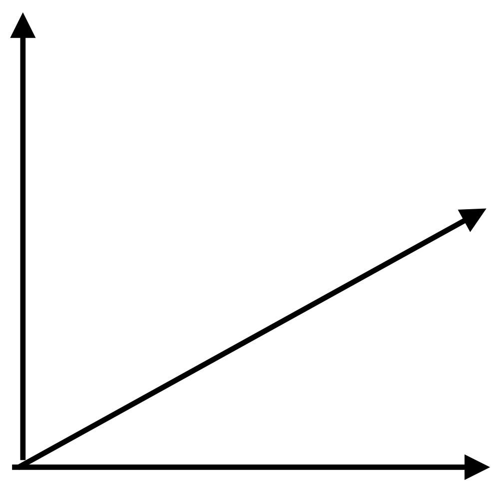
Note: both height and weight are input of a datum, so we don't have a label y for each datum.

high-dimensional space of face images

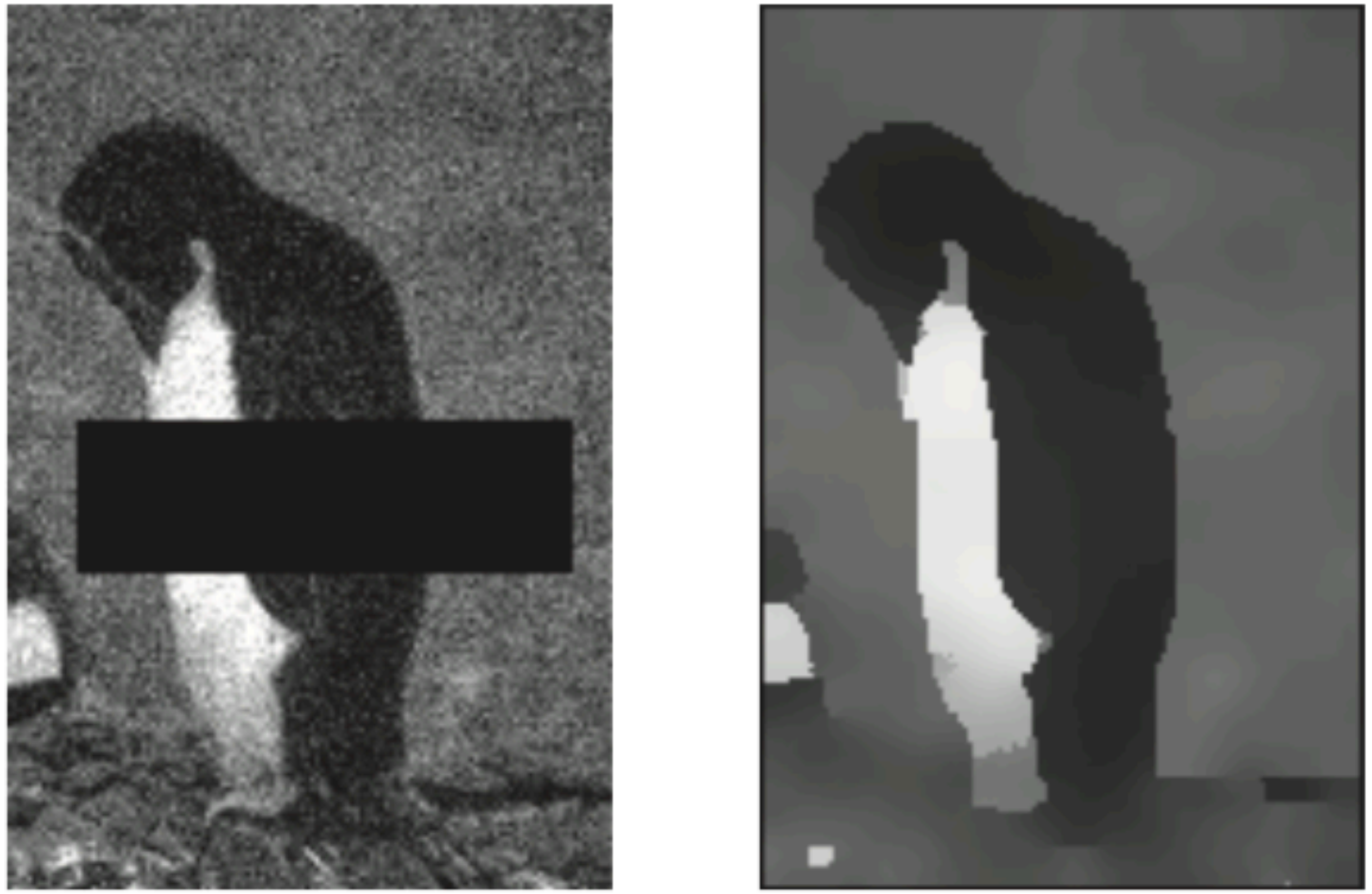
each image is a point in 64 x 64 = 4096 dimensional space

principle-component analysis (PCA)

A face image can be thought of as a vector in high dimensional space. The important features are the projected component onto the principal direction vectors.



missing pixels completion

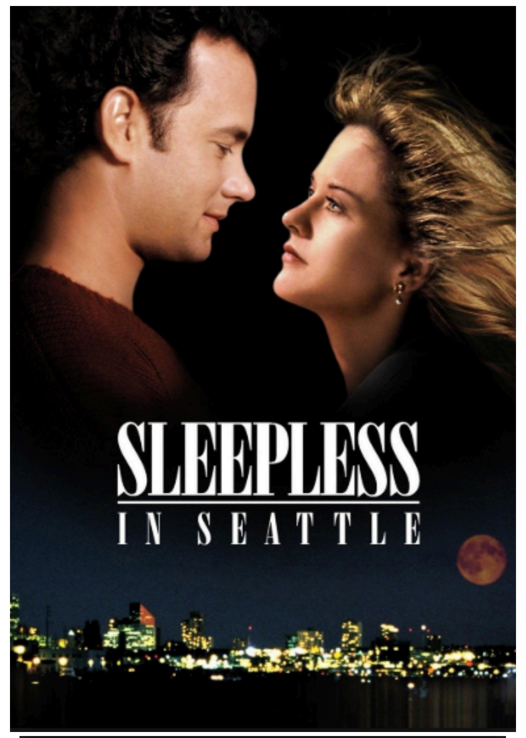


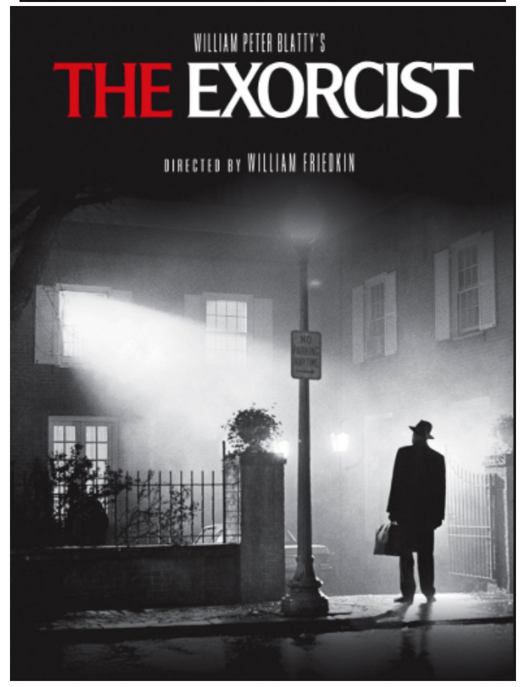
(a) (b)

pixels reconstruction from
statistics of test images

read 1.3.4.1

recommender system,
collaborative filtering

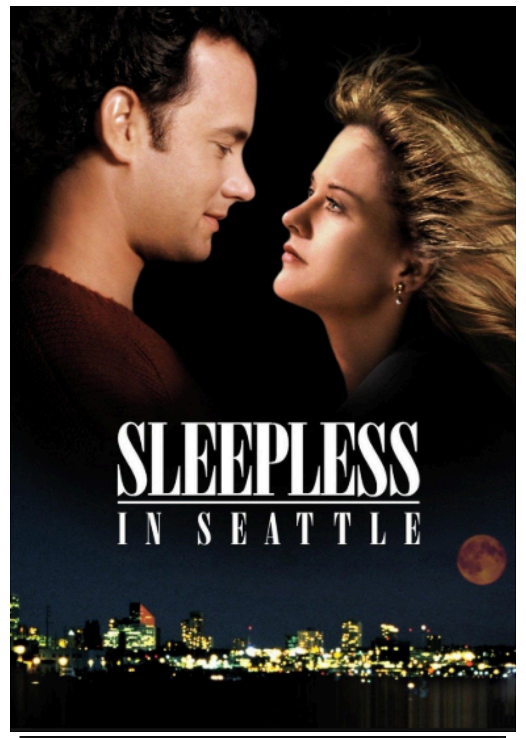


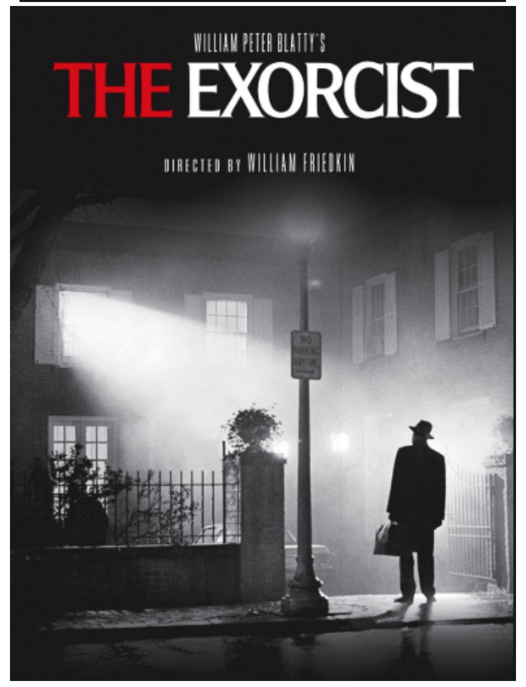


1		?	3	5	?
?	1				2
	4		4	5	?

think about how YouTube generates
the recommending videos to you.

recommender system,
collaborative filtering





users

movies

1		?	3	5	?
?	1				2
	4		4	5	?

think about how YouTube generates
the recommending videos to you.

Natural language processing

Why the free services from FB, google, twitter is actually profitable for them ...

Data set: e-mails from all google users, wikipedia articles, twitter comments, Ig photos, etc...

Huge neural network

When I _____ my ______ ______ ____________, ____

Natural language processing

Why the free services from FB, google, twitter is actually profitable for them ...

Data set: e-mails from all google users, wikipedia articles, twitter comments, Ig photos, etc...

↓

Huge neural network

When I visit my ______ ______ __________, ______________________________

Natural language processing

Why the free services from FB, google, twitter is actually profitable for them ...

Data set: e-mails from all google users, wikipedia articles, twitter comments, Ig photos, etc...

↓

Huge neural network

When I ______ my _______ _______ _______________, ____________________________________
visit grandma in nursing house

Natural language processing

Why the free services from FB, google, twitter is actually profitable for them ...

Data set: e-mails from all google users, wikipedia articles, twitter comments, Ig photos, etc...

↓

Huge neural network

When I visit my grandma in nursing house, I would like to buy a flower for her

Natural language processing

Why the free services from FB, google, twitter is actually profitable for them ...

Data set: e-mails from all google users, wikipedia articles, twitter comments, Ig photos, etc...

↓

Huge neural network

When I visit my grandma in nursing house, I would like to buy a flower for her

call

Natural language processing

Why the free services from FB, google, twitter is actually profitable for them ...

Data set: e-mails from all google users, wikipedia articles, twitter comments, Ig photos, etc...

↓

Huge neural network

visit grandma in nursing house I would like to buy a flower for her

When I ______ my ______ ______ ____________, ______________________________

call my brother in Army

Natural language processing

Why the free services from FB, google, twitter is actually profitable for them ...

Data set: e-mails from all google users, wikipedia articles, twitter comments, Ig photos, etc...

↓

Huge neural network

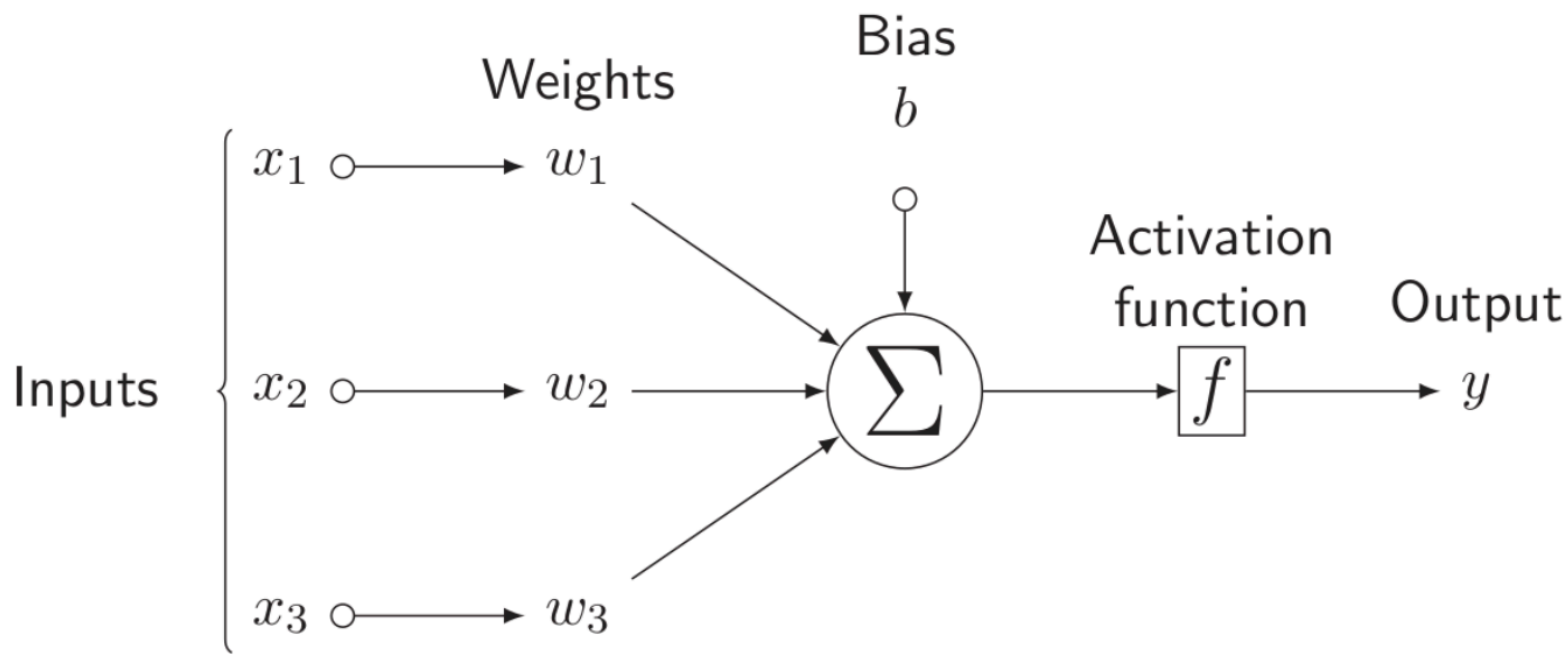
visit grandma in nursing house I would like to buy a flower for her

When I ______ my ______ ______ ____________, ____________________________

call my brother in Army I want to tell him the flight fare is good for next two weeks.

Breaking down neural networks

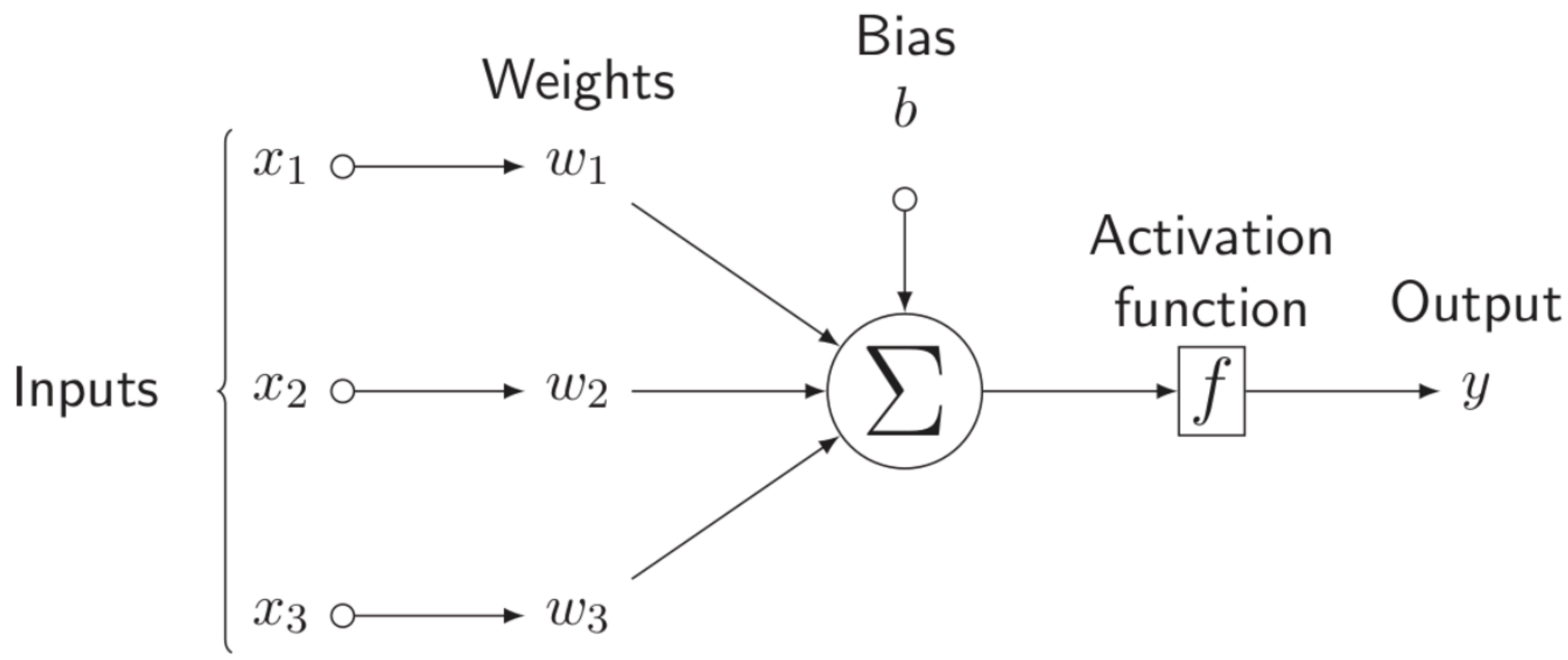
Regression (week 3 & 4) and logistic regression (week 5 &6)



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

Breaking down neural networks

Regression (week 3 & 4) and logistic regression (week 5 &6)

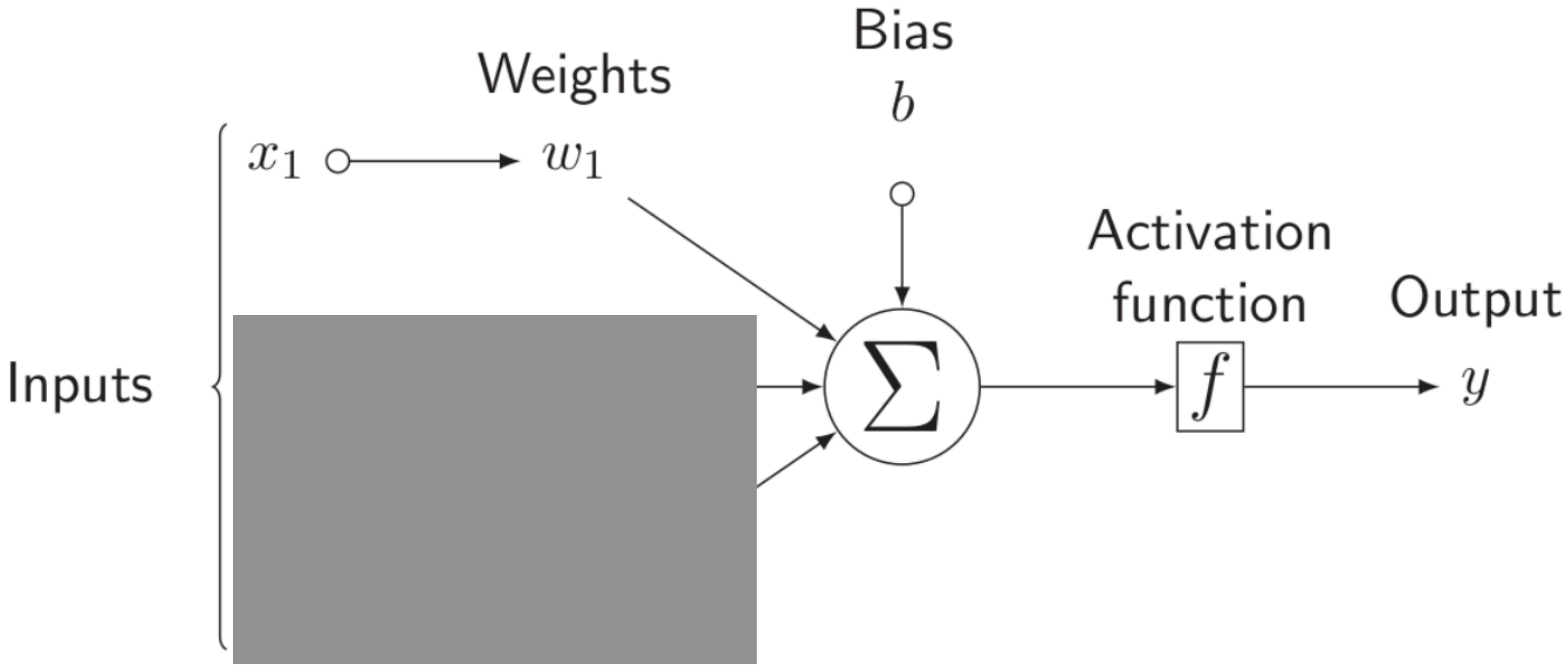


https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network

Regression:
$\mathbf{x} \in \mathbb{R}^d$ and $y \in \mathbb{R}$

Breaking down neural networks

Regression (week 3 & 4) and logistic regression (week 5 &6)

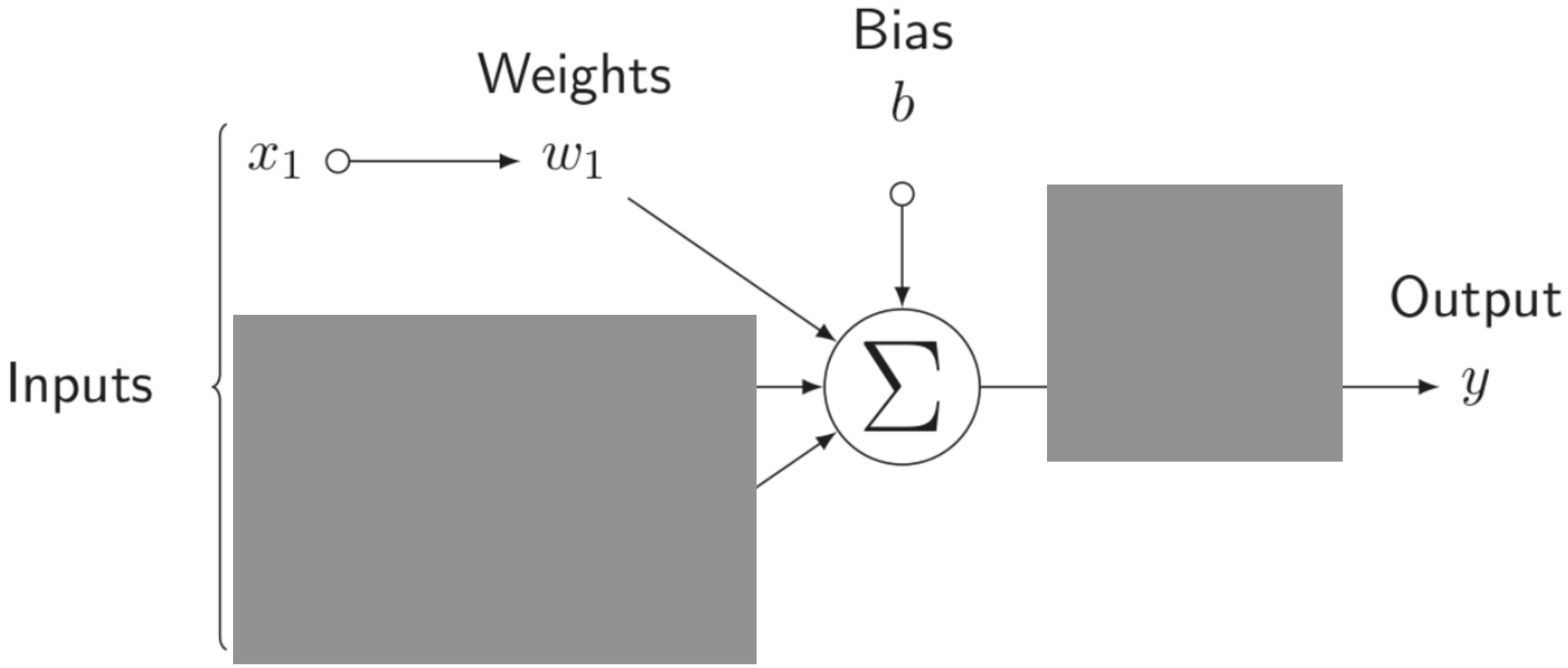


https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network

Regression:
$\mathbf{x} \in \mathbb{R}^d$ and $y \in \mathbb{R}$

Breaking down neural networks

Regression (week 3 & 4) and logistic regression (week 5 &6)

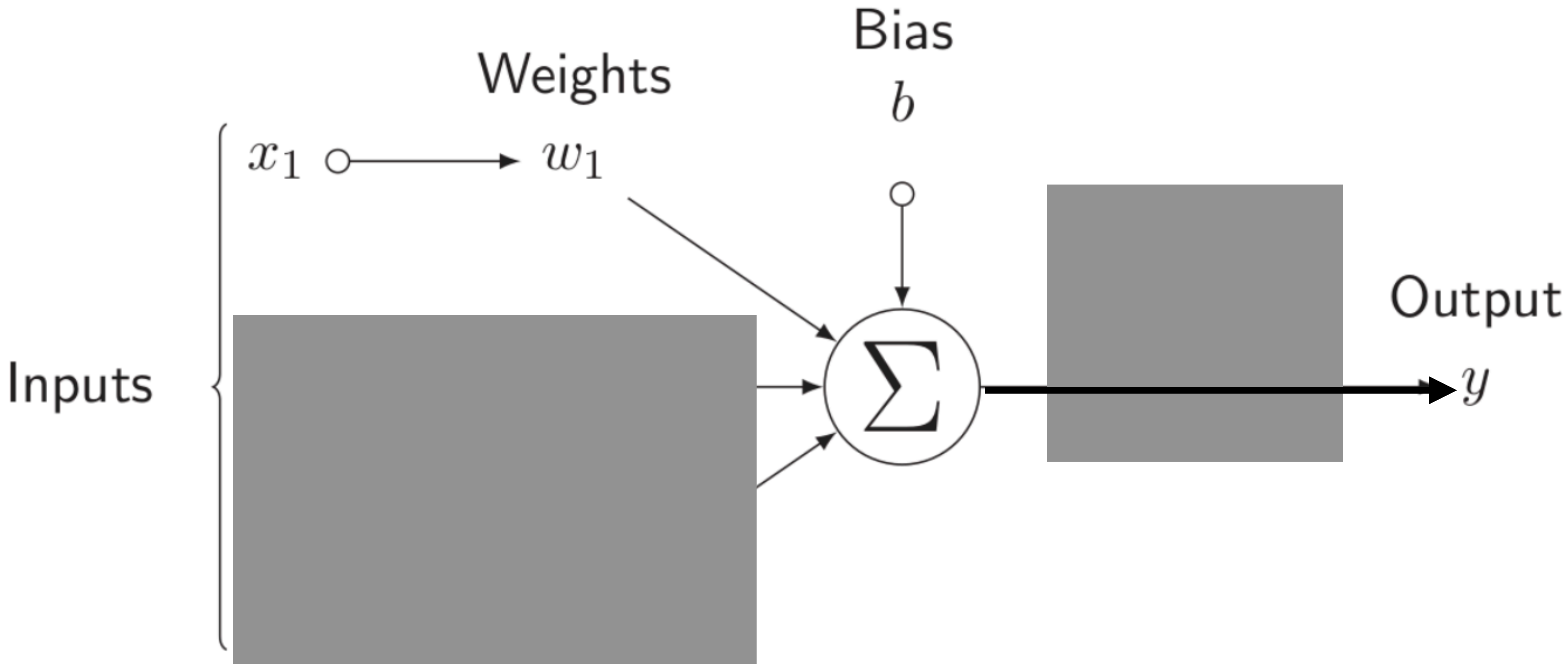


https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network

Regression:
$\mathbf{x} \in \mathbb{R}^d$ and $y \in \mathbb{R}$

Breaking down neural networks

Regression (week 3 & 4) and logistic regression (week 5 &6)

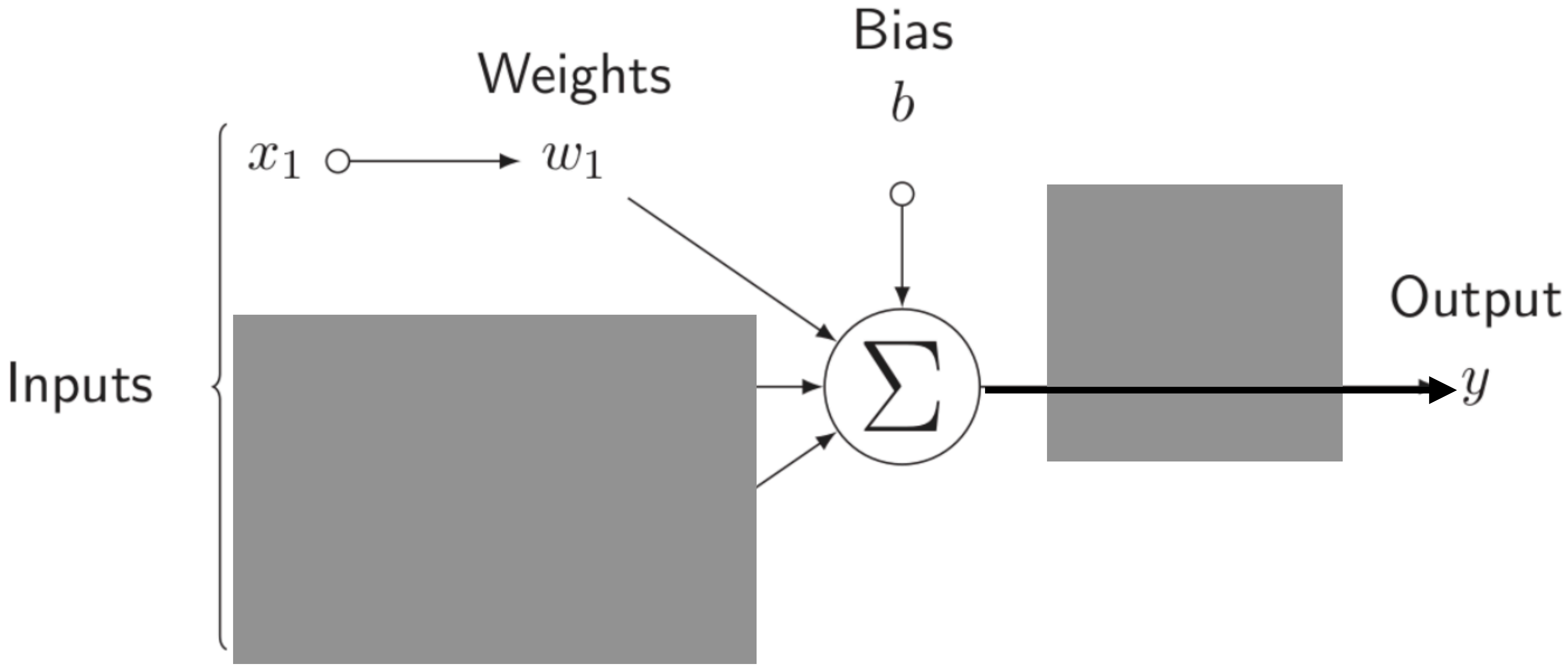


https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network

Regression:
$\mathbf{x} \in \mathbb{R}^d$ and $y \in \mathbb{R}$

Breaking down neural networks

Regression (week 3 & 4) and logistic regression (week 5 &6)



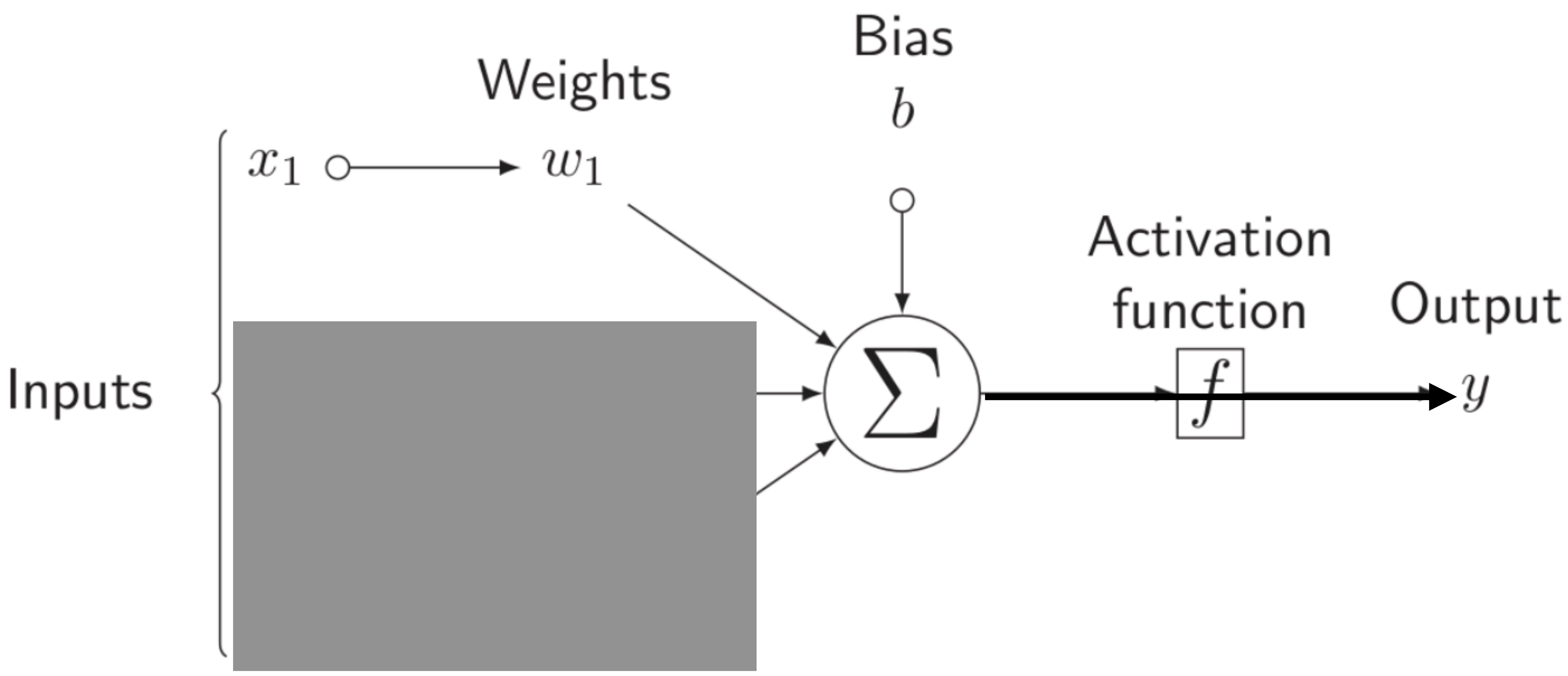
https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network

Regression:
$\mathbf{x} \in \mathbb{R}^d$ and $y \in \mathbb{R}$

Logistic Regression:
$\mathbf{x} \in \mathbb{R}^d$ and $y \in \{0,1\}$

Breaking down neural networks

Regression (week 3 & 4) and logistic regression (week 5 &6)



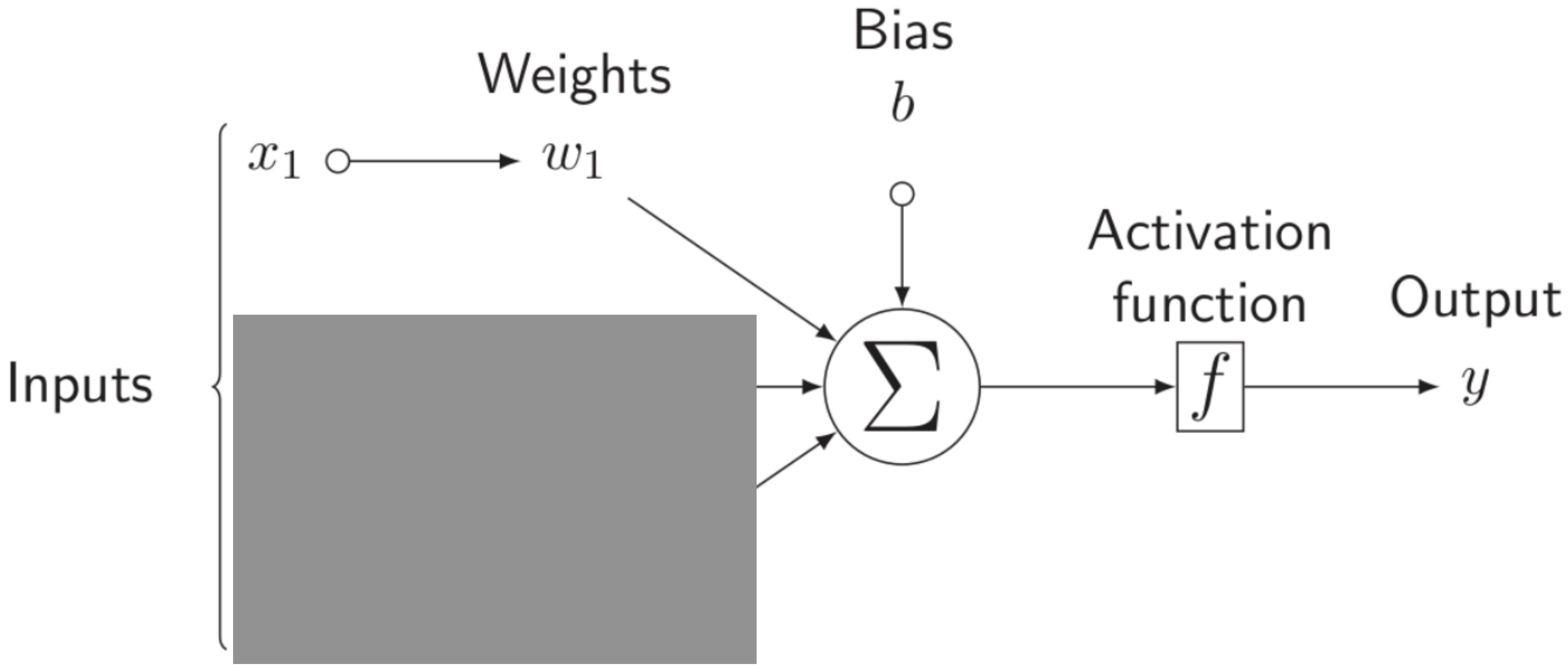
https://tex.stackexchange.com/questions/132444/diagram-of-an-artificial-neural-network

Regression:
$\mathbf{x} \in \mathbb{R}^d$ and $y \in \mathbb{R}$

Logistic Regression:
$\mathbf{x} \in \mathbb{R}^d$ and $y \in \{0,1\}$

Breaking down neural networks

Regression (week 3 & 4) and logistic regression (week 5 &6)



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

Regression:
$\mathbf{x} \in \mathbb{R}^d$ and $y \in \mathbb{R}$

Logistic Regression:
$\mathbf{x} \in \mathbb{R}^d$ and $y \in \{0,1\}$