U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Introduction to Deep Learning using Keras

FPRO 2018/2019

Pedro M. Ferreira

INESC TEC, PDEEC

10 January 2019

What is Machine Learning?

What is Machine Learning?

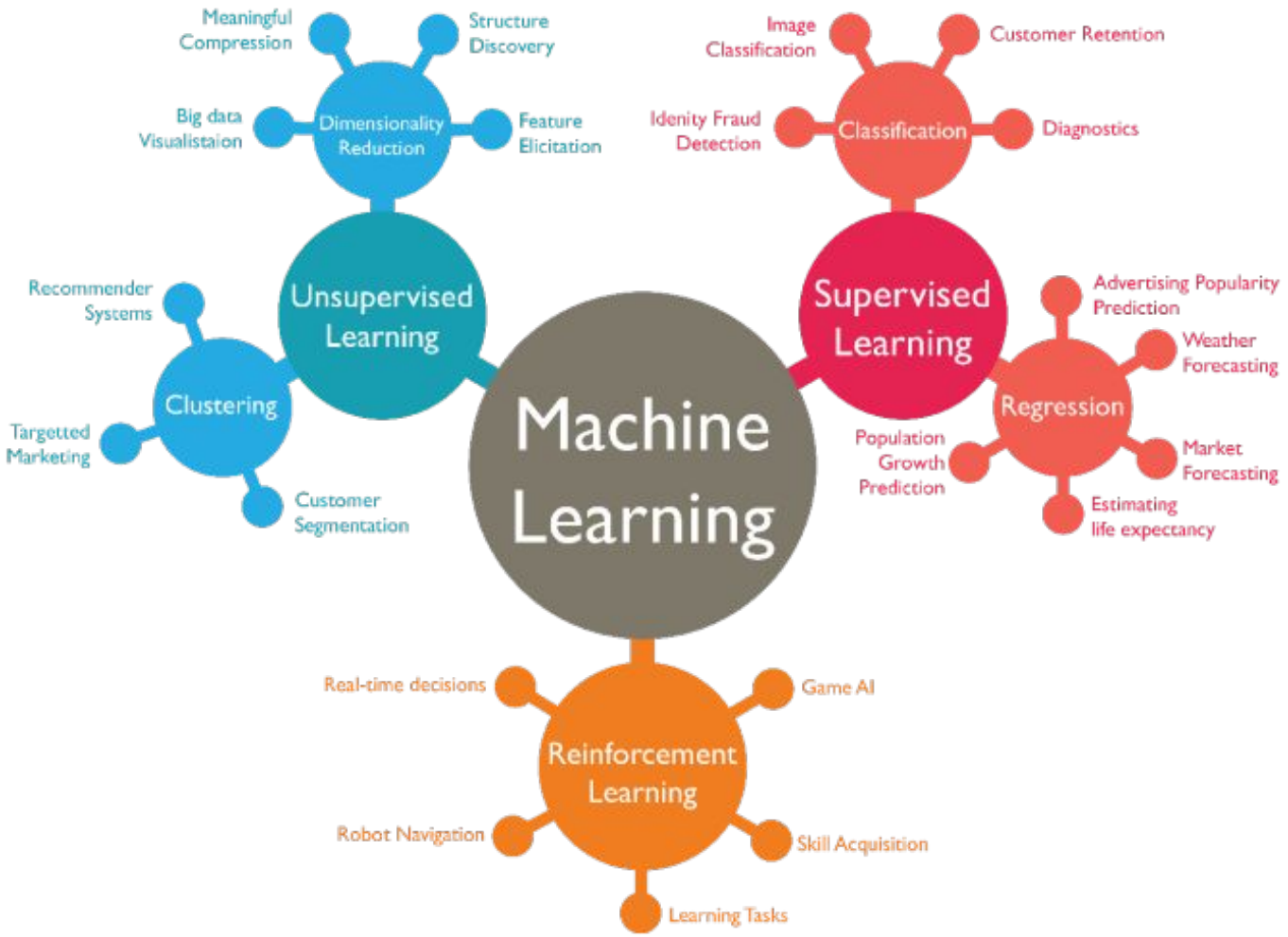
by Arthur Samuel, 1959

"the field of study that gives computers the ability to learn without being explicitly programmed."

by Tom M. Mitchell, 1997

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

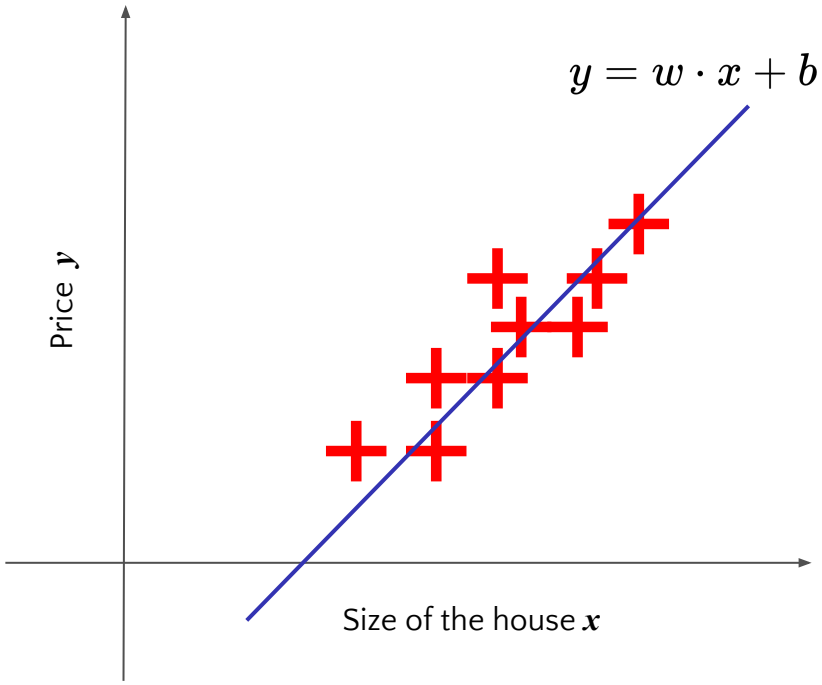
Machine Learning Problems



Deep Learning

What is a Neural Network?

[EXAMPLE] Housing Price Prediction

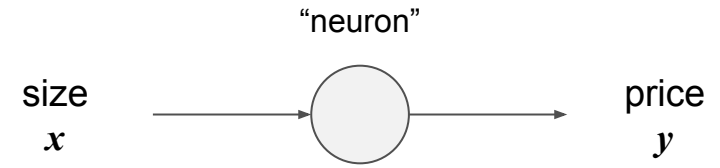


› Using **linear regression**, you can fit a linear function to predict the price y as a function of the size of the house x:

$$y = w \cdot x + b$$

where w and b are learned from the data.

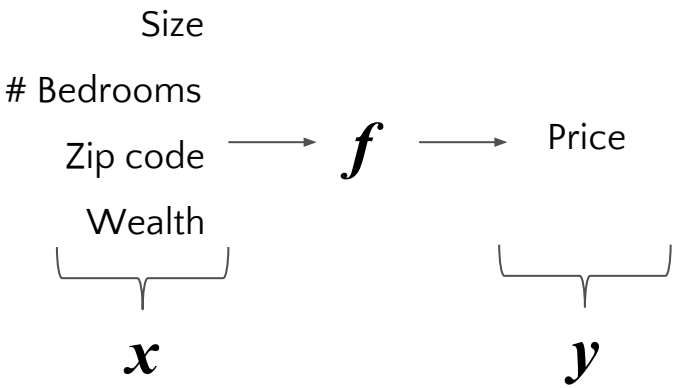
› You can think of this function as a **very simple neural network**!



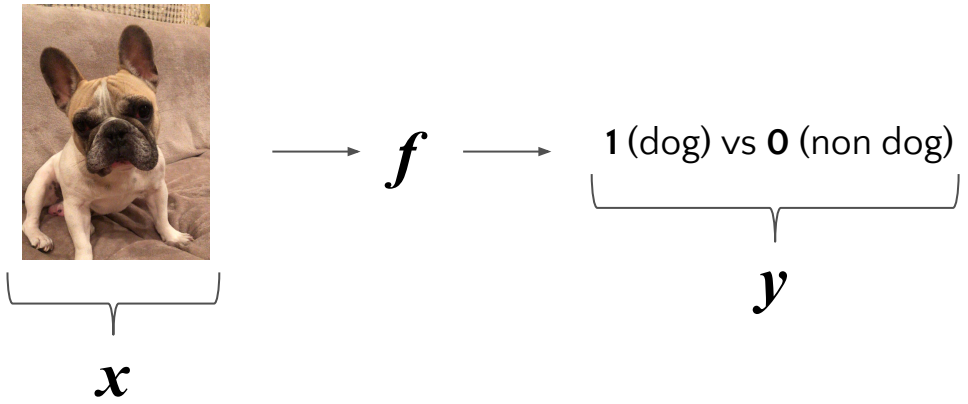
Simple neural network with one hidden unit/neuron.

What is a Neural Network?

So, what type of functions f should we fit to the data when we don't know the relation/structure of the data?



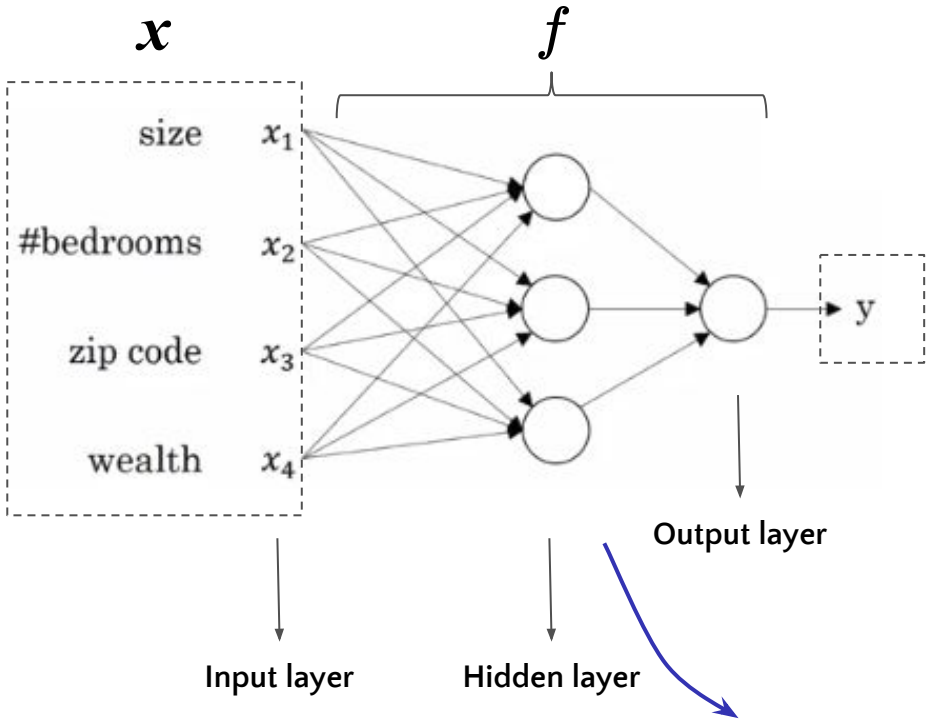
Housing Price Prediction



Binary image classification

What is a Neural Network?

So, what type of functions f should we fit to the data when we don't know the relation/structure of the data?



› Given enough training examples with both x (observations) and y (labels), **neural networks are remarkably good at figuring out functions f that accurately map from x to y** !

$$f : x \longrightarrow y$$

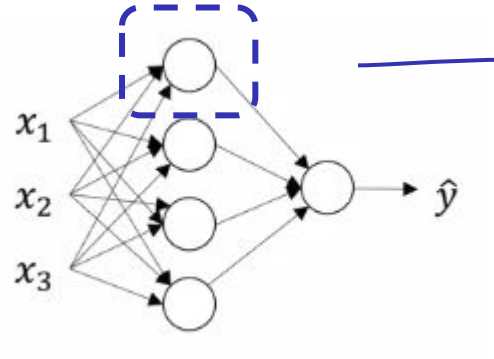
› Building neural networks is analogous to Lego bricks:

we take individual neurons and stack them together to create complex neural networks.

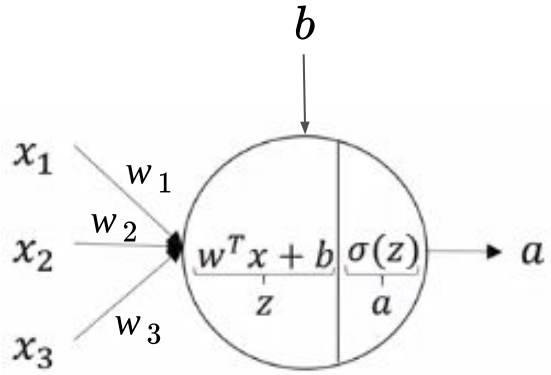
We talk in **deep learning** when we have a **neural net** with **two or more hidden layers**!

What is a Neural Network?

Neural Network Representation



Output of a single neuron



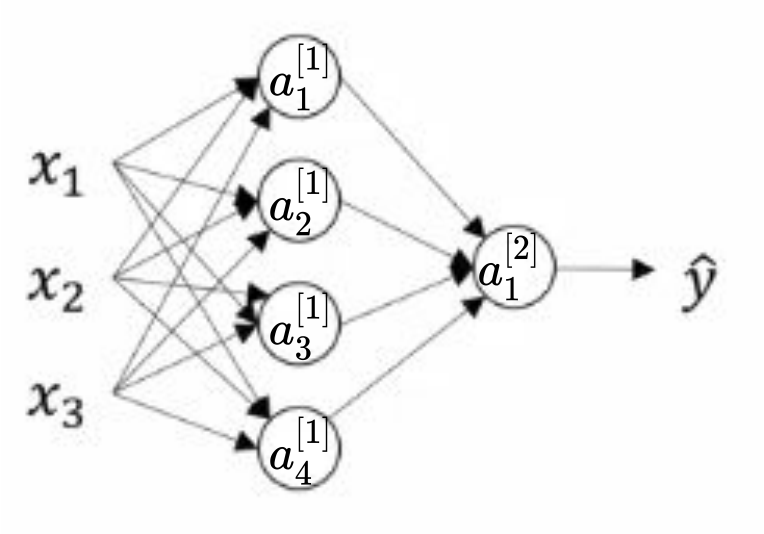
$$z = w^T x + b$$
$$a = \sigma(z)$$

$$w^\top = [w_1, w_2, w_3], x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Activation function/Non-linearity

What is a Neural Network?

Neural Network Representation



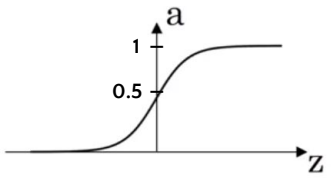
› Given X: $z^{[l]}$ and $a^{[l]}$ of layer l are given by:

$$z^{[1]} = W^{[1]}x + b^{[1]}$$
$$a^{[1]} = \sigma(z^{[1]})$$
$$z^{[2]} = W^{[2]}a^{[1]} + b^{[2]}$$
$$a^{[2]} = \sigma(z^{[2]})$$

$\hat{y} = a^{[2]}$ ⟶ Predicted output

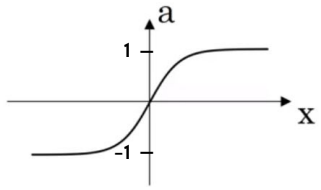
› Why non-linear activation functions are important?

Sigmoid



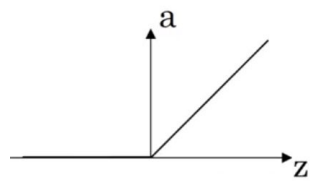
$$a = \frac{1}{1+e^{-z}}$$

Tanh



$$a = \frac{e^{z}-e^{-z}}{e^{z}+e^{z}}$$

ReLU

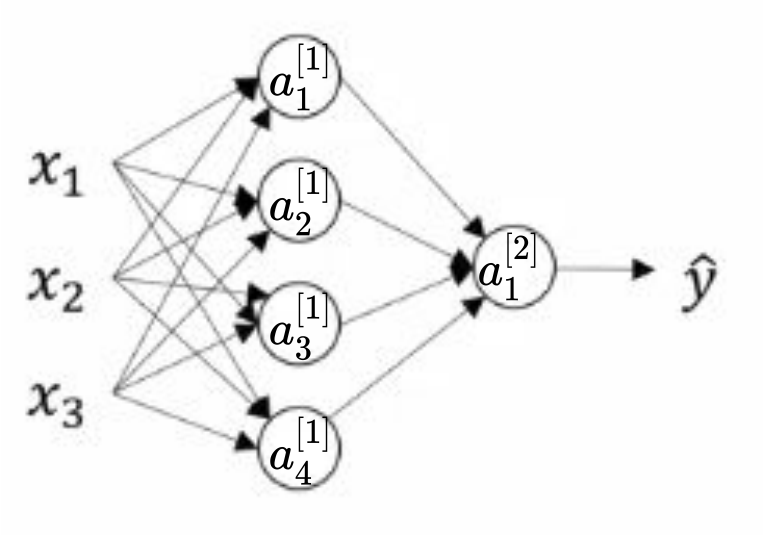


$$a = \max(0, z)$$

- Because the composition of two linear functions is still a linear function!
- Unless you throw a non-linear there, you are not computing more interesting functions as you go deeper in the network!

What is a Neural Network?

Neural Network Representation



› Given X: $z^{[l]}$ and $a^{[l]}$ of layer l are given by:

$$z^{[1]} = W^{[1]}x + b^{[1]}$$
$$a^{[1]} = \sigma(z^{[1]})$$
$$z^{[2]} = W^{[2]}a^{[1]} + b^{[2]}$$
$$a^{[2]} = \sigma(z^{[2]})$$

$\hat{y} = a^{[2]}$ → **Predicted output**

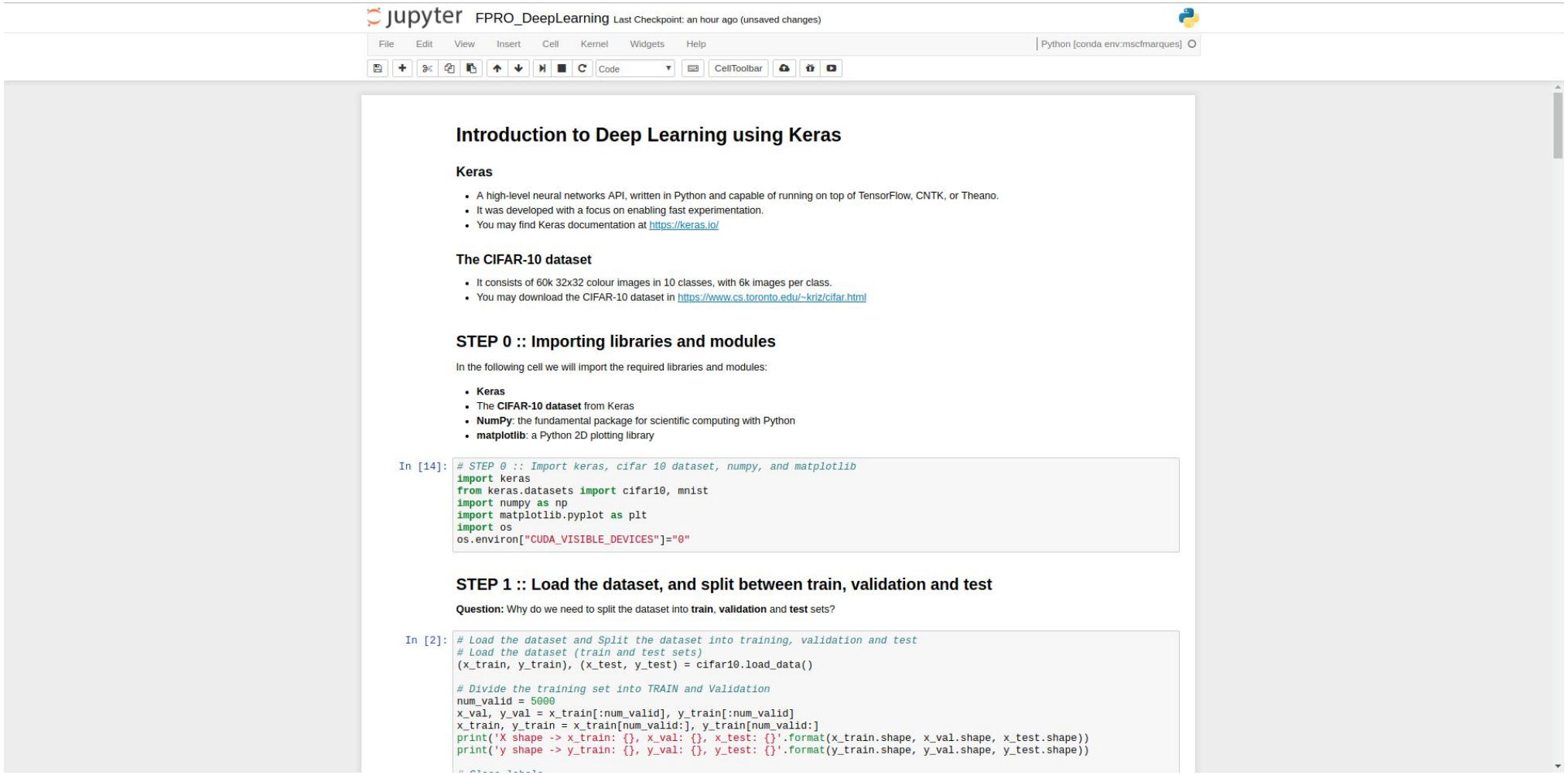
How can we learn the best possible parameters W, so that the function learned by the neural network accurately maps from x to y?

⇩

This is called the **training/fitting process** of the neural network!

Implementing Neural Nets using Keras

Implementing Neural Networks



EXERCISES

- Moodle activity at: LE26: Deep Learning