

# Training

We had left out the area of training until now.

The network is fed with sample data for training.

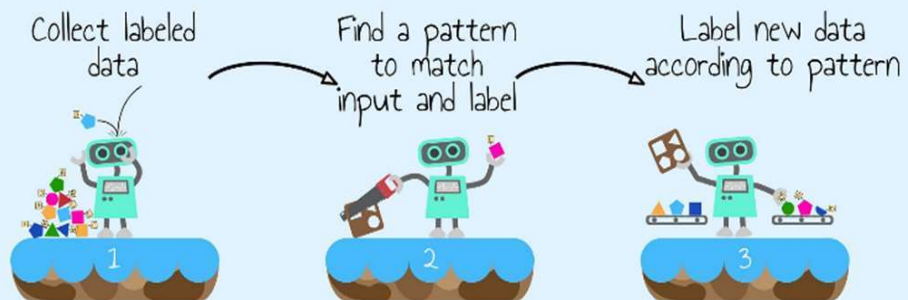
```
["f", 60, 165]    # Alice  
["m", 73, 183]    # Bob  
["m", 69, 178]    # Charlie  
["f", 54, 152]    # Diana
```

The data are marked with "m" (male) or "f" (female).  
This is called supervised learning.

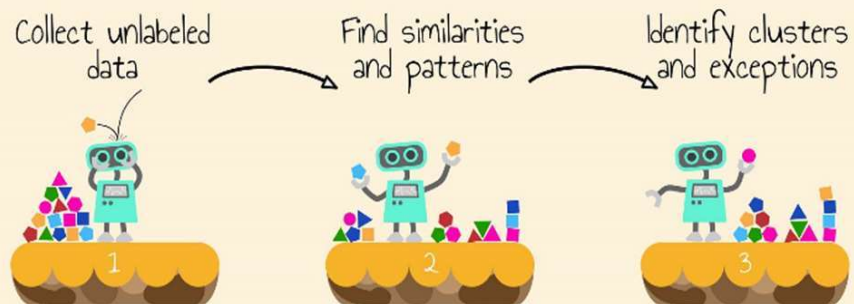
Overall, we can distinguish:

## Types of Machine Learning

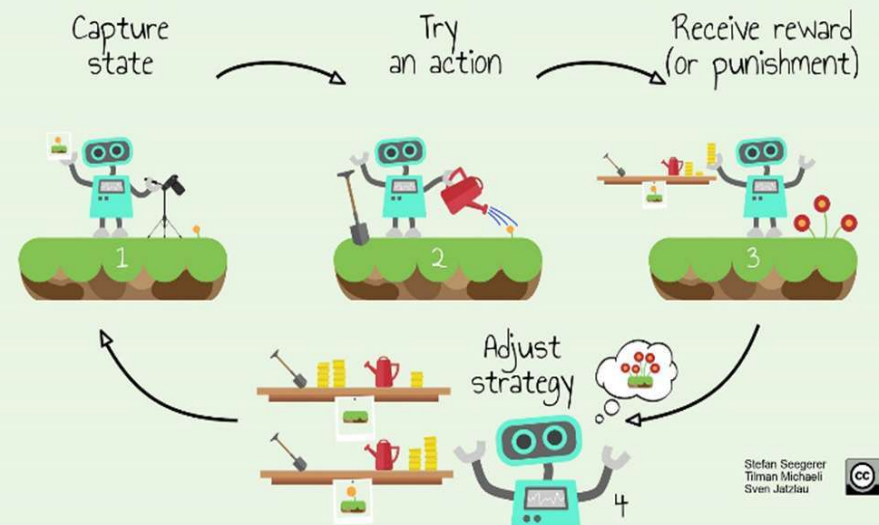
### Supervised Learning



### Unsupervised Learning



### Reinforcement Learning



# Implementation

The technical implementation, we want to explain here only very briefly and not in detail:

Initially all the weights and biases are assigned randomly.  
An error value can be assigned to each such combination.

# How to calculate the error value

The mean squared error (MSE) is often used for this.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

MSE = mean squared error

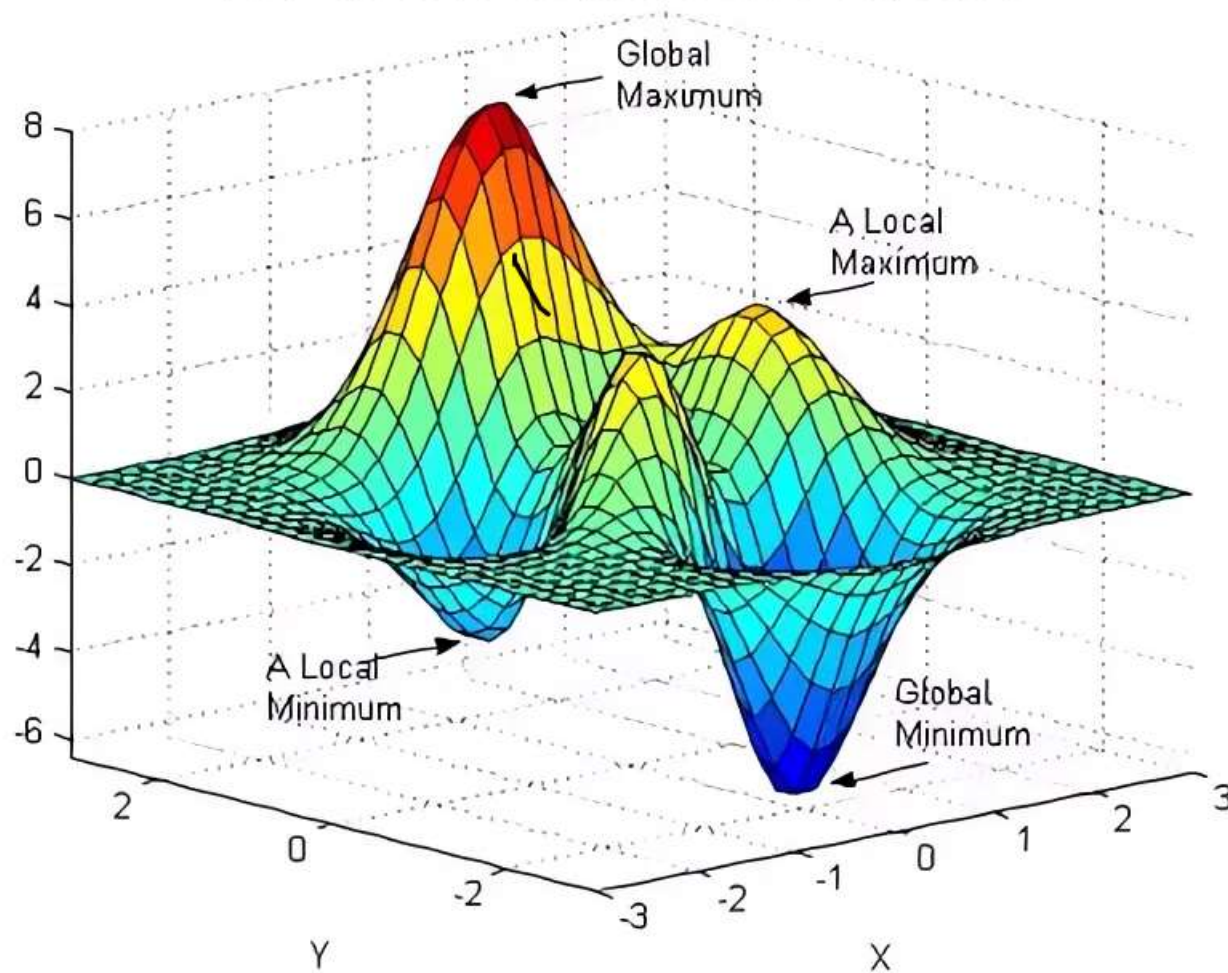
$n$  = number of data points

$Y_i$  = observed values

$\hat{Y}_i$  = predicted values

$n$  observations were made. In each case, the difference between the expected value (" $f$ "=1, " $m$ "=0) and the observed value is squared. Then all values are summed and the sum is divided by  $n$ .

A Non-Convex Combination of Gaussian Distributions



Zachary kaplan, CC BY-SA 4.0, via Wikimedia Commons

Together with the error value, the parameters span an  $n$ -dimensional plane. To illustrate, here is the case for only 2 parameters. The search for a solution here corresponds to the search for (local or global) minima in this plane.

# Gradient Descent

Very often, but not exclusively, the so-called "gradient descent method" is used here. It determines the slope at the current point (gradient) and tries to move downhill towards the minimum.

The network was trained and successfully tested in this way. The resulting weights and biases can then be read out. Try to do the calculation yourself for the 2 pairs of values given.

Is it a man or a woman?

You can also try a value yourself.

# Notes

Before the calculation, the average values for weight and height are subtracted from the pair of values. This is how the network was trained. The sigmoid function is used as the constraint function, see the [sigmoid.pdf](#) file below.