# Foundations of Machine Learning

It’s of fundamental importance to use the correct terminology!

In machine learning we have no hypothesis on the data -> ML finds the function explaining the relationship between the features and the continuous value of a class

## Definitions

**Pattern**

* Numerical pattern
  + Values associated with measurable and ordinal **features**
  + Continuous or discrete
  + Represented as numerical vectors in multidimensional space
  + e.g. counts of a detector, area of a spectrum, signal intensity…
* Categorical pattern
  + Values associated with qualitative features or the presence/absence of a feature
  + Not convertible to ordered numerical values
  + e.g. gender, nationality, …
* Sequential patterns
  + With spatial or temporal relationship *(features)*
  + Fixed or variable sequences length
  + e.g. sequence of image frames (video), sequence of sounds (stream audio)
  + Images have spatial relationship and signals have temporal relationship
* Other structured data
  + Patterns organized in complex structures
  + e.g. genomic data, language (for natural language processing)

**Task**

* **Classification** (classification / detection)
  + **Class** = set of **patterns** with *common properties*
  + Training a function capable of performing the mapping from the pattern space to the class space
    - e.g. patients with the disease or without the disease is a class -> the common property is “with disease” or “without disease”
    - e.g. students are either male or female. Male is the common property of class “male” and female is the common property of class “female”.
  + Assigning (predicting) a *pattern* to a *class*
    - 2 classes -> binary classification
    - More than 2 classes -> multi-class classification
  + Pattern space: dimension of the pattern
  + Class space: dimension of the class
* **Regression**
  + Training a function capable of *finding* the *relationship between features* *and* the *continuous values of a class*
  + Assigning (predicting) a continuous value to a numerical pattern
  + Not to be confused with interpolation (but similar as a task)
    - e.g. energy of a power plant, dosage of a drug administration, time of disease relapse
* **Clustering**
  + Identifying groups (*clusters*) of patterns with similar features
  + The identified clusters can be used as classes (further grouped if necessary)+
  + The number of clusters may not be known in advance

## Task in the field of vision

Pattern = object in an image (static or dynamic)

A pattern has a set of features

* Classification
  + Associates the class with the object
* Localization
  + Determines the position of the object in a bounding box
* Detection
  + Associates the class and determines the position of multiple objects (in bounding boxes)
* Segmentation
  + Associates individual pixels of objects with different classes (in coloured pixels and determines the position in contours) -> each object is highlighted and classified
* Image quality improvement
  + Improves the quality of an image
* Improved image reconstruction
  + Completes/speeds up the reconstruction

## Learning methods

* Supervised -> ONLY ONE USED FOR HEALTHCARE
  + The classes of patterns used for learning are known a priori
* Unsupervised
  + The classes of patterns used for learning are NOT known a priori
* Semi-supervised

## Algorithm parameters

The learning behaviour of an algorithm of ML depends on a series of ***parameters***

* e.g. weights of a neural network

Learning is reached optimizing these parameters according to an ***objective function***

* when we converge to a value for the objective function

## Objective function

It can be:

* optimization function of the algorithm (e.g. Euclidean distance)
* a performance function (metric of performance: accuracy, sensibility or specificity)
  + -> maximise it
* a error/loss/cost function
  + -> minimise it

## Training, validation, testing set

* Training set is the set of patterns on which the learning happens
* Validation set is the set of patterns on which **hyperparameters** are cross-validated
* Testing set is the set of patterns on which to evaluate the actual performances of the model.
  + This dataset cannot contain patterns used neither for training nor validation

## Hyperparameters

Hyperparameters are parameters whose values control the learning process and determine the values of model parameters that a learning algorithm ends up learning.

* The prefix ‘hyper\_’ suggests that they are ‘top-level’ parameters that control the learning process and the model parameters that result from it.

Hyperparameters are used by the learning algorithm when it is learning but they are not part of the resulting model.

At the end of the learning process, we have the trained model parameters which effectively is what we refer to as the model.

The hyperparameters that were used during training are not part of this model.

* Train-test split ratio
* Learning rate in optimization algorithms (e.g. gradient descent)
* Choice of optimization algorithm (e.g., gradient descent, stochastic gradient descent, or Adam optimizer)
* Choice of activation function in a neural network (NN) layer (e.g. Sigmoid, ReLU, Tanh)
* The choice of cost or loss function the model will use
* Number of hidden layers in a NN
* Number of activation units in each layer
* The drop-out rate in NN (dropout probability)
* Number of iterations (epochs) in training a NN
* Number of clusters in a clustering task
* Kernel or filter size in convolutional layers
* Pooling size
* Batch size

## Parameters

Parameters on the other hand are internal to the model. That is, they are learned or estimated purely from the data during training as the algorithm used tries to learn the mapping between the input features and the labels or targets.

Model training typically starts with parameters being initialized to some values (random values or set to zeros). As training/learning progresses the initial values are updated using an optimization algorithm (e.g. gradient descent). The learning algorithm is continuously updating the parameter values as learning progress but hyperparameter values set by the model designer remain unchanged.

At the end of the learning process, model parameters are what constitute the model itself.

* The coefficients (or weights) of linear and logistic regression models.
* Weights and biases of a NN
* The cluster centroids in clustering

## Partitioning training, validation and testing set

70-15-15