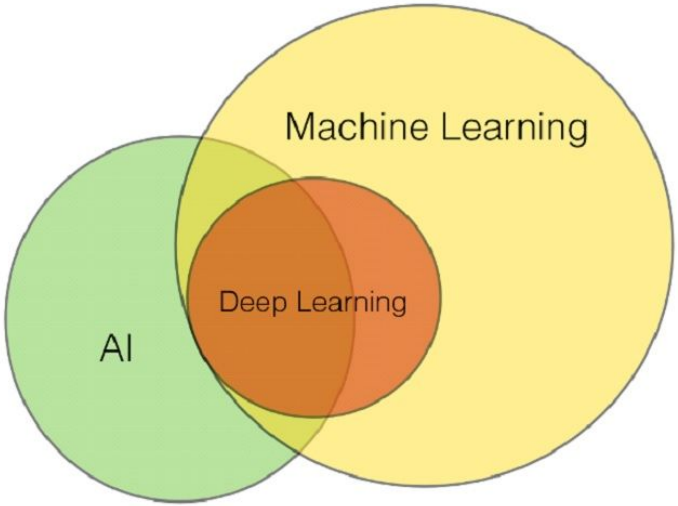
Introduction to Machine Learning and AI

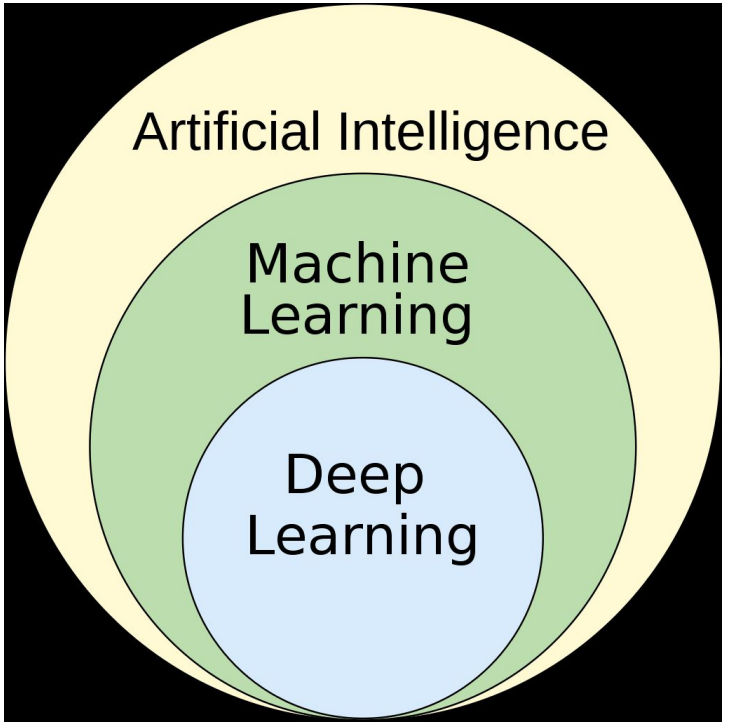
Silvan David Peter - Emmanouil Karystinaios

Kunstuniversität Linz
Linz University of Arts

Institute of Computational Perception

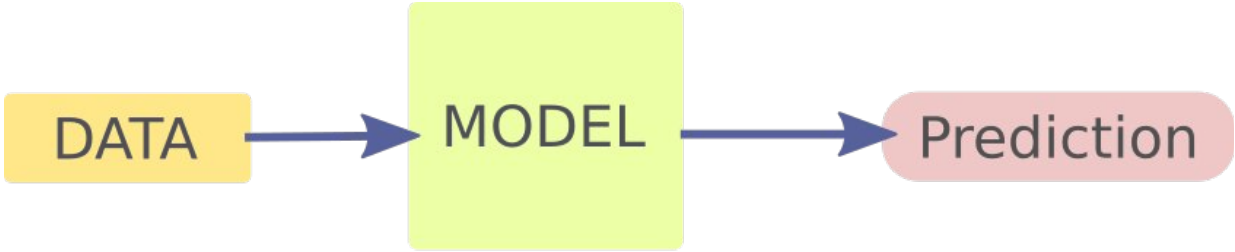
What is AI?


Machine Learning
Deep Learning
AI


Artificial Intelligence
Machine Learning
Deep Learning

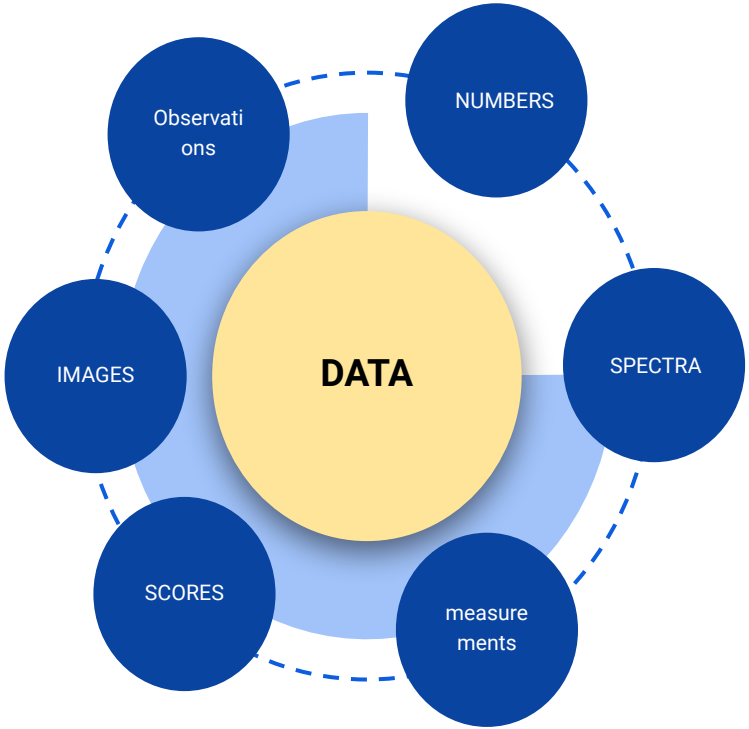
Learning from Data

- What is Data?
- What are Predictions?
- What is Model?



What is Data?

Data is information.
What matters in Learning systems is that this information is represented numerically as a Number, Vector, Matrix, or tensor.



Predictions

Prediction is information we can extract from data with the help of some model.

For example in Image recognition the data is an image of a dog, then the prediction would be the **label** *“dog”***.**

In systems these labels are represented by numbers, for example:

1 = *dog*, 2 = *cat*, 3 = *car*, etc..

Image from Wikipedia

Trainable Model

A model simply refers to some real-valued function for example:

$$f(x) = x^2 + 1$$

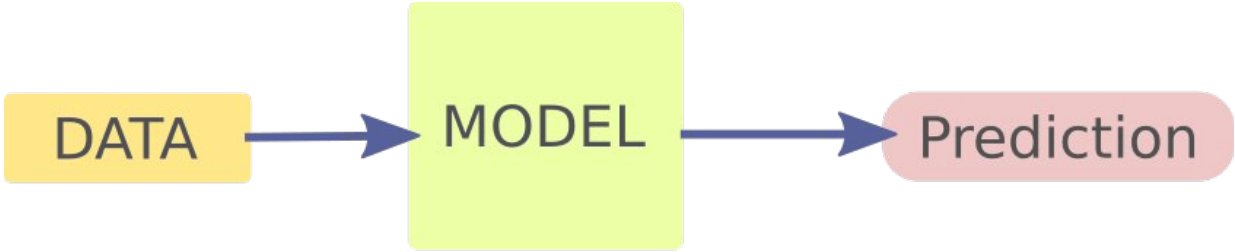
A function is trainable if we can change the value of its parameters:

$$f(x) = ax^2 + b$$

In this case a, b

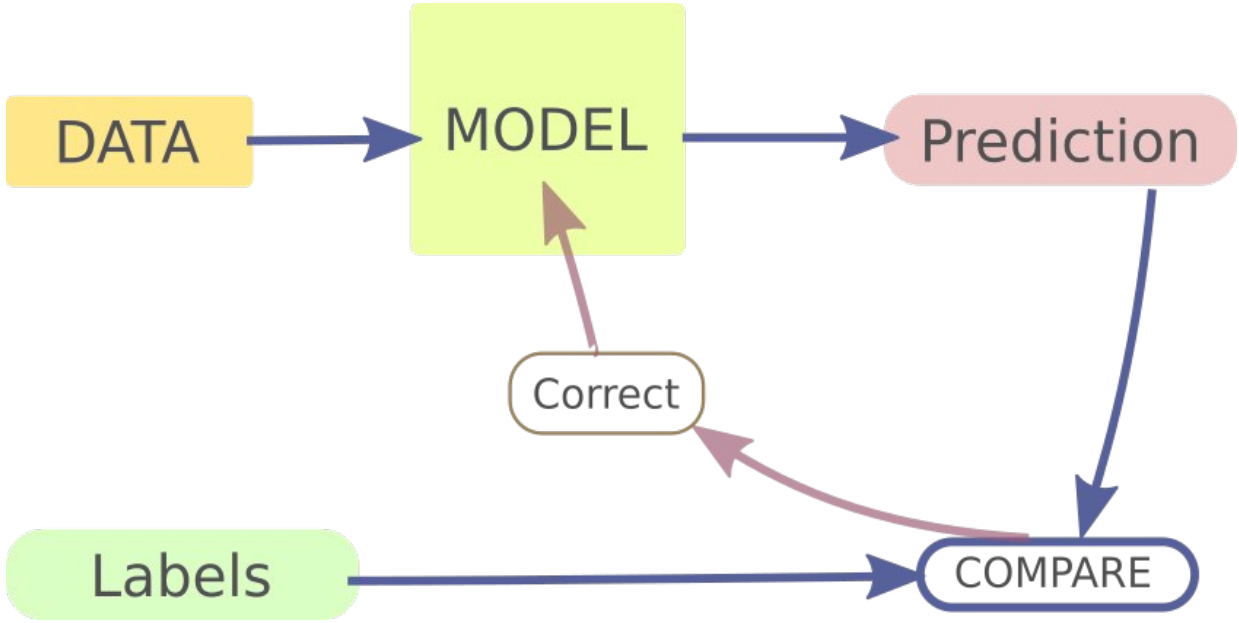
Learning from Data

- What is Data?
  Data is Information
- What are Predictions?
  Information extracted from data with the help of a model.
- What is Model?
  Some kind of processing of data.



Supervised Learning

- What are Labels?
- How do we compare?
- How do we correct?



Terminology Summary

- Label = the variable we are predicting
- Features = the input variables describing our data
- Example/Sample = a particular instance of data, i.e. x
  - A labeled example has (features, label), i.e. (x, y)
  - A unlabeled examples has no label (used for making predictions on new data)
- Model = maps examples to predicted labels y’

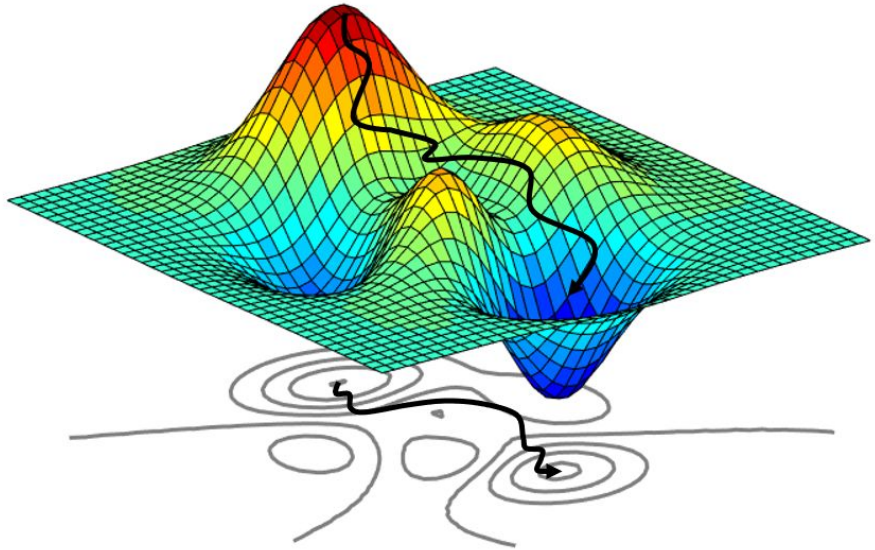
Loss Function

Loss function is a method of evaluating how well your algorithm models your dataset. If your predictions are totally off, your loss function will output a higher number. If they're pretty good, it'll output a lower number.

Optimization

In the simplest case, an optimization problem consists of maximizing or minimizing a real function by systematically choosing input values from within an allowed set and computing the value of the function.

A standard algorithm for optimizing models in Machine Learning is called **Gradient Descent.** The most widely used is called **ADAM**.



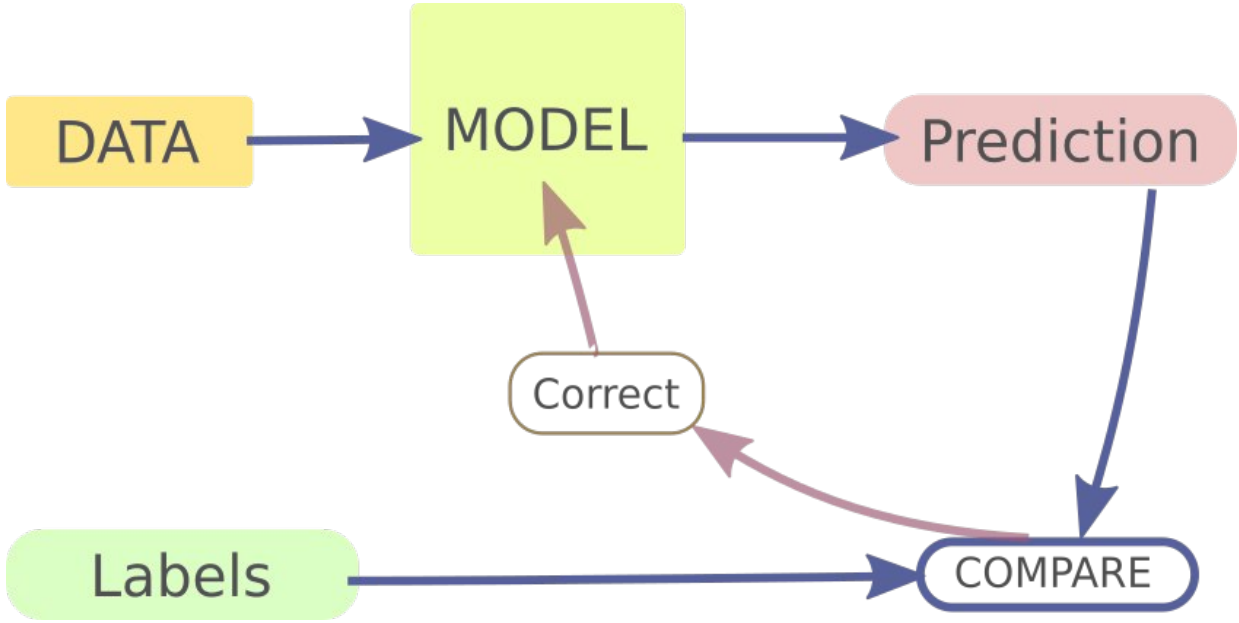
Supervised Learning

- What are Labels?
- How do we compare?

  Loss Function

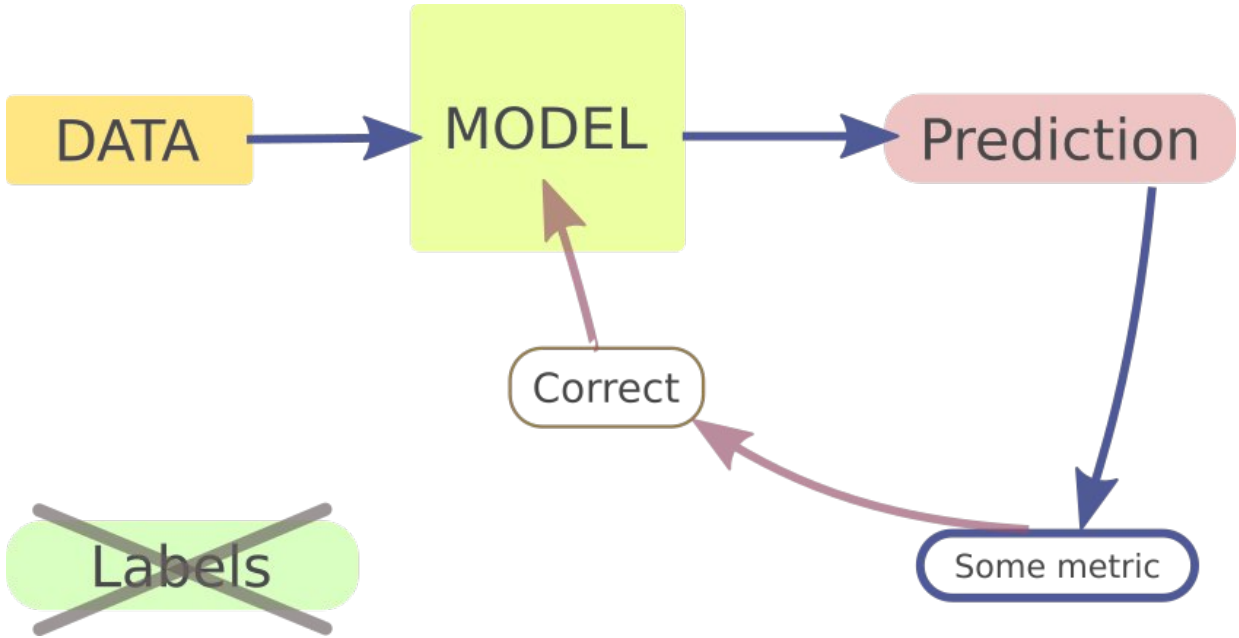
- How do we correct?

  Optimization (ADAM)



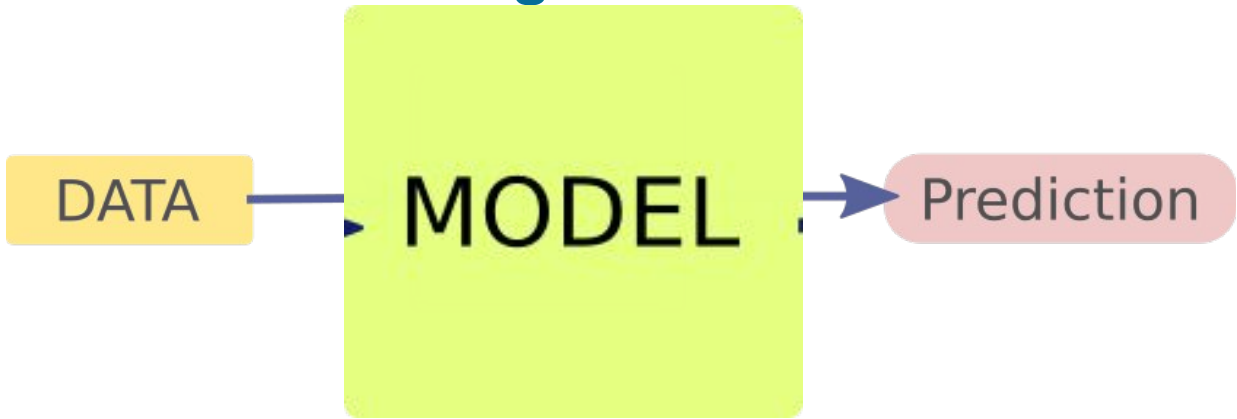
Unsupervised Learning

- How do we compare without Labels?
- Examples?



What is a Model in Machine Learning

- Scalar Multiplication
- Linear Transformation
- Non-Linear Transformation
- Neural Network



Regression vs Classification

A **regression** model predicts continuous values. For example, regression models make predictions that answer questions like the following:

- What is the value of the Spotify stock?
- What is expected volume of the next track?

A **classification** model predicts discrete values. For example, classification models make predictions that answer questions like the following:

- Given a song will the user skip or not skip?
- Is this the genre of this song of a rock, pop or classical?

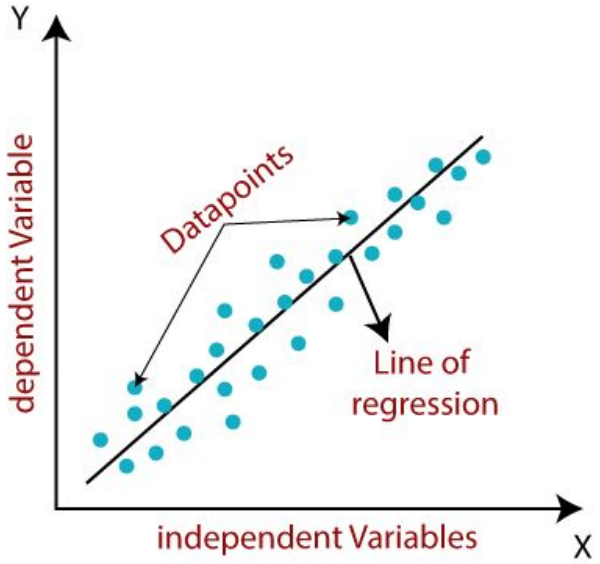
Linear Regression - One dimension

Linear Regression is used to model the relationship between two variables and estimate the value of a response by using a line-of-best-fit.

The variables in question (x,y) here are assumed to have a linear relation.
The question we are asking in this scenario is:

Given *f(x) = ax + b* which values *a, b* are more suitable such as *f(x) ≈ y.*

$$y_{pred} = ax + b$$



Linear Transformation

When x and/or y are not just numbers, but vectors or matrices, then the previous example works exactly the same.

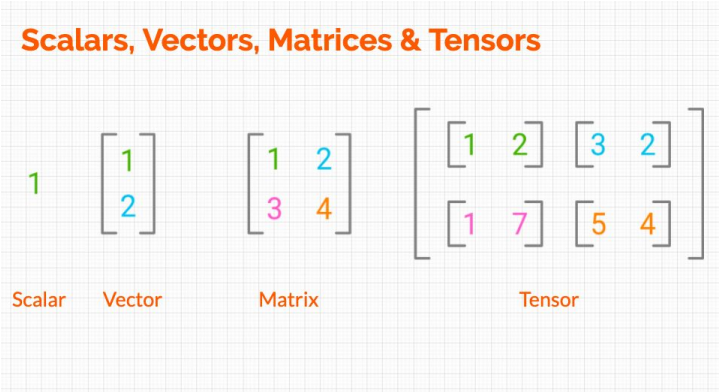
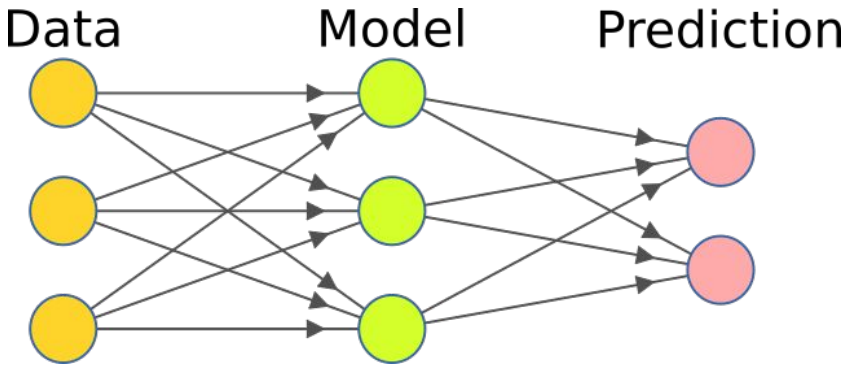


Image from Towardsdatascience.org



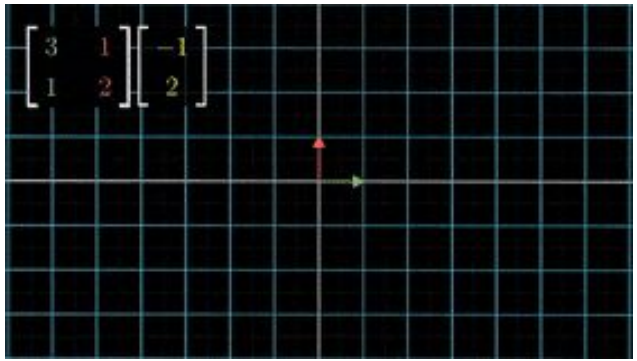


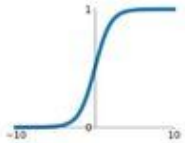
Image from 3Blue1Brown

Non-Linear Operations

Activation Functions

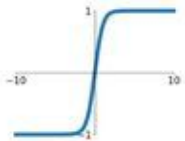
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



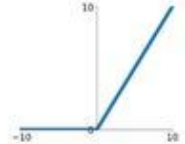
tanh

$$\tanh(x)$$



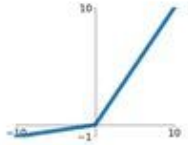
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

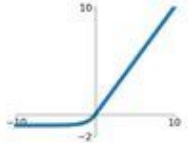


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

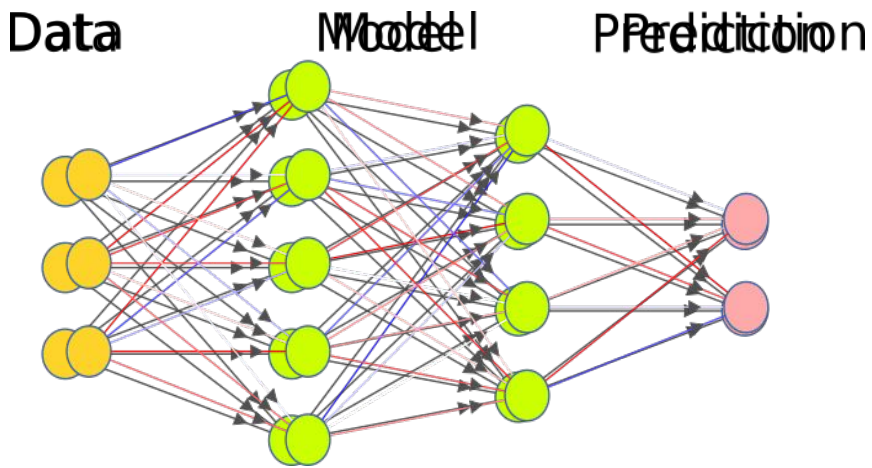
$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



that take as input the result

Neural Network - Weights and Biases

Weight is **the parameter within a neural network that transforms input data within the network's hidden layers**. A neural network is a series of nodes, or neurons. Within each node is a set of inputs, weight, and a bias value.



Stages of Learning

TRAINING

During training we update the parameters of a model.

The training process is divided in two hierarchical sections:

1. Epochs (Number of circles over the training data)
2. Batches (Iterating over subsets of the training data)

VALIDATION/TESTING

During validation or testing we check if the model parameters after a session of training result to a good prediction. In other words we evaluate the model without changing its parameters.

PREDICTION

The prediction phase is the same as the validation phase but in this case we do not have a way of evaluating. After a model is trained, we feed it data for which we don't necessarily have labels and we get a prediction.

External Resources

- https://www.youtube.com/watch?v=aircAruvnKk&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&ab_channel=3Blue1Brown
  - Youtube Series for mathematical background of Neural Networks (3Blue1Brown)
- https://scikit-learn.org/
  - Introductory Python Package for Machine Learning.