Artificial Intelligence

KS Im Lee

Kantonsschule Im Lee

Banuijan Yogananthan

Maturarbeit

Contents

0 Introduction

This text-paper discusses the fundamentals of artificial intelligence, which includes processes like machine learning, deep learning and their mathematical/geometrical background. Also, we will take a closer look at different types of so called "neural networks" and deduce the advantages/disadvantages of the important ones. It is meant to be used as an introduction to a Matura project dealing with the construction of AI's. By the end, one should receive a brief overview about artificial intelligence.

1 What is Artificial Intelligence

AI, or artificial intelligence, is prevalent in our world. Its use is versatile: voice/text recognition such as Siri or Alexa, finance, video games, military, art, and even the government make use of this technology. With AI, processes that could take ages when done by a human can be completed within seconds by implementing elements of AI. That could be a definition of AI: It's a human, but just better. AI is basically exceeding or matching the capabilities of a human. So, it tries to match the intelligence, whatever that means, and capabilities of a human subject.

There are different abilities most AI's have in common :

- ability to discover
- ability to infer
- read information
- ability to reason
- ability to figure things out

To create an artificial intelligence, different types of frameworks can be used. These frameworks serve as a model through which we can load our data. In the following pages, we will discuss how and which methods are used for the construction of such systems.

2 Machine Learning

The term "Machine Learning" was introduced in 1959 by the computer scientist "Arthur Samuel". It's a subset of AI and contains these two fundamental abilities:

- classification of data
- accurate predictions based on data

How it works:

In order to train our model, we need to collect data which will be fed to it. A diagram of the data could be represented in a coordinate system:

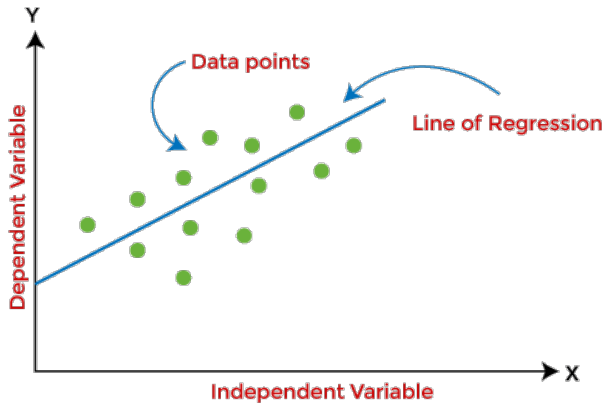


Figure 1: Data Points visualized

There are different factors which could influence the quality of the model. First of all, the data point should contain all the necessary information. If not, the accuracy will decrease. The same is for the "varianz". If the "varianz" is high, there will be a change in the "line of regression". For high variances, more data is needed to get an accurate output. Other methods like normalising, converting or randomising values are also used for preparing data sets, which we won't go in detail. The same data should not occur twice.

After preparing the raw data, we need to come up with techniques which we could train our model with.

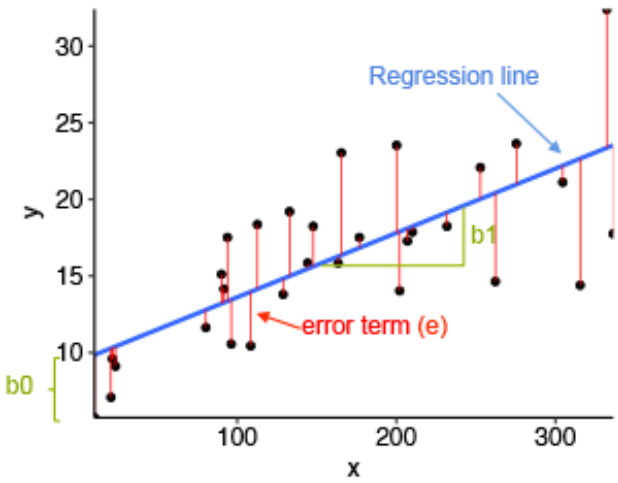
Supervised Learning

In supervised learning, the model will be trained with labelled data. Labelling here means that the size which one wants to preach is contained in the data. It is not enough to put in the data and let it train. It needs certain information about the object. Therefore, the connection in the data is already given before.

Again, supervised learning is based on 2 concepts

Regression

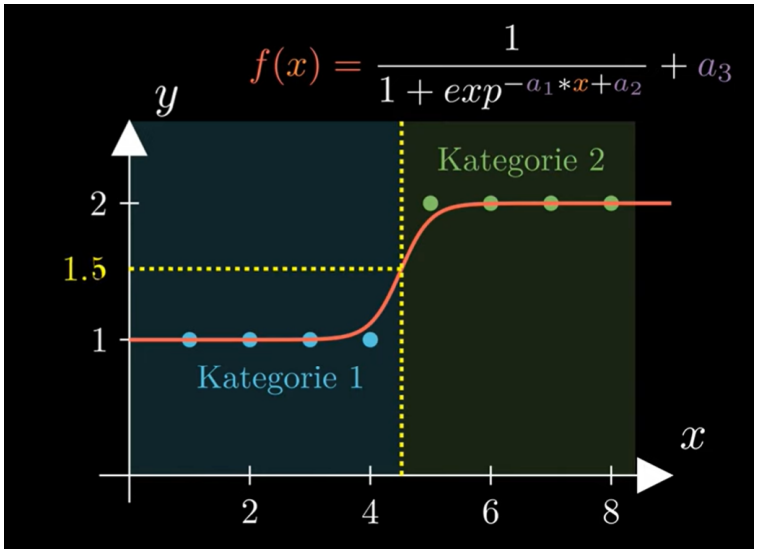
Consider a regression line: $f(x) = mx + q$. The line in this case will be determined by minimizing the distance from the line to all the data points by adjusting the parameter m and q:



$$min \sum_{i=0}^{N} d_i^2 \tag{1}$$

Classification

consider following function : $f(x) = \frac{1}{1+exp^{-a_1*x+a_2}} + a_3$:



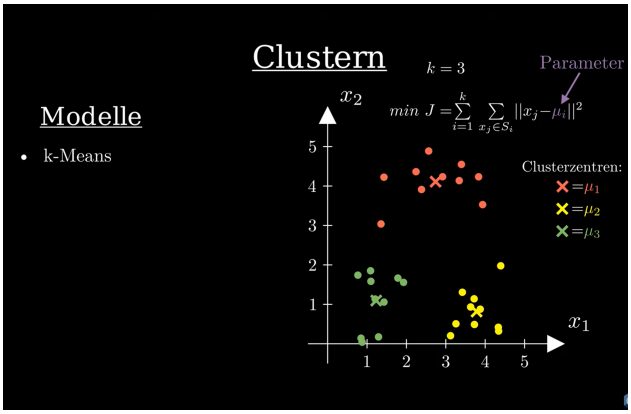
As you can see, depending on your output, it determines whether it belongs to category 1 or 2. This technique is used to classify data and is also part of supervised learning.

Unsupervised Learning

In contrast to supervised learning, this learning technique uses unlabelled data. This means, the preached size in the dataset is not given already before. The connection of the data is determined by the algorithm by its own.

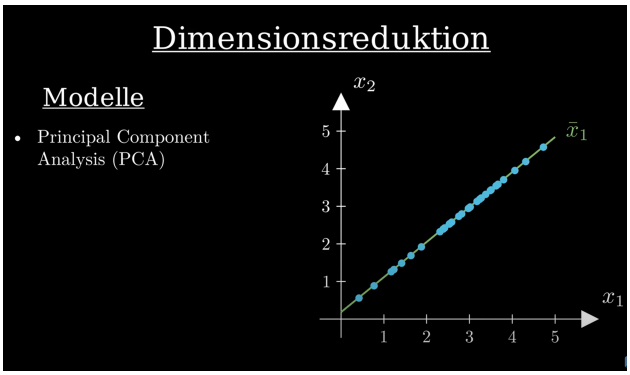
The idea of unsupervised learning is again based on 2 concepts :

Clustering



Clustering is a concept where data points are categorized through so-called cluster-centres (by formula shown above). One could imagine these cluster-centres to span a circle where all the respective points are included. In contrast to the concept of classification where data is basically assigned to two "connection types", clustering creates the connection and the data is respectively assigned.

Reduction of Dimensions



This concepts yields (as the name tells us) a reduction of dimensions. This is not achieved by leaving out dimensions, it rather depends on the goals which are aimed onto (e.g reducing information loss). This is done by transforming the base of the coordinate system so that the points lie on one straight line

3 Deep Learning

Deep learning is a subsequent field of machine learning. In fact, deep learning is also based on the same concepts we have discussed above. The major difference of ML and DL are the amount of layers a neural network consists of. While ML - neural networks consist of 3 layers (Input-, hidden-, Output layer), deep learning networks consist of more than 3 layers. Machine learning is rather used for simpler concepts (e.g categorization of sth.). Deep learning is more used with more complex ideas, e.g language recognition.

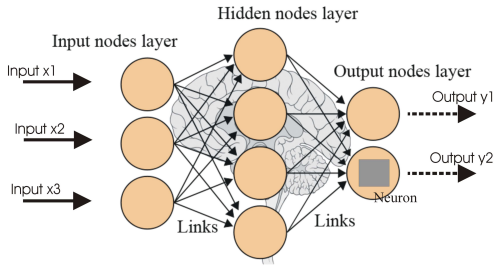


Figure 2: ML NN

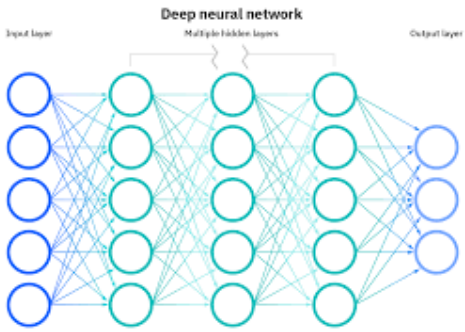


Figure 3: DL NN

Neural Networks

As this concept is the basis of deep learning, we will focus on the concept of a neural net and try to understand how it works.

How it works

In order to understand the mechanism which lies behind Neural Networks, we will focus on understanding the individual components, namely neurons and weights. We will use a "Plain Vanilla" model, since it's the easiest one.

Neurons

Neurons are placeholders for specific values. These values are called "activation". In a way, you can imagine that if this value is high, the neurons will be more intensive, meaning this value has a high importance.

Weights

Weights are used to connect the layers of a neural network. With higher values, you can determine the importance of each neuron. As we dig deeper in examples, one can observe how certain patterns can be determined by setting up specific weight values

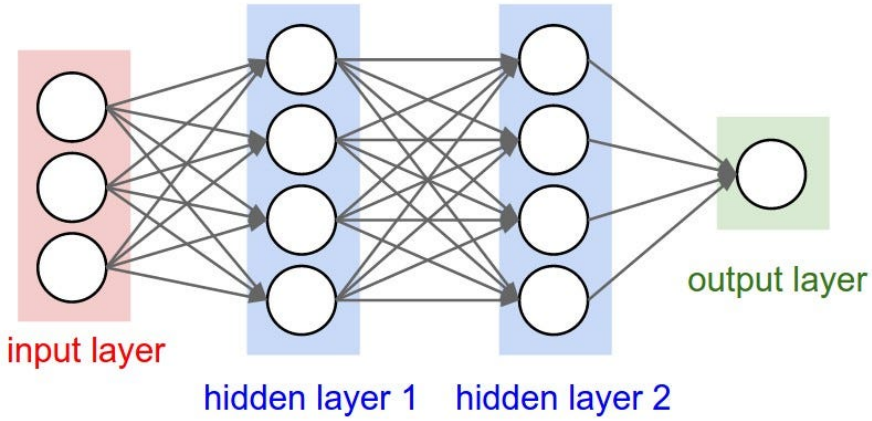
bias

A bias is a number you add after taking the scalar of all weight and "activations" to make sure the output is higher than the minimum

activation function

An activation function is used because normally, you want your values to be between 0 and 1. A common function which would to this is the sigmoid function.

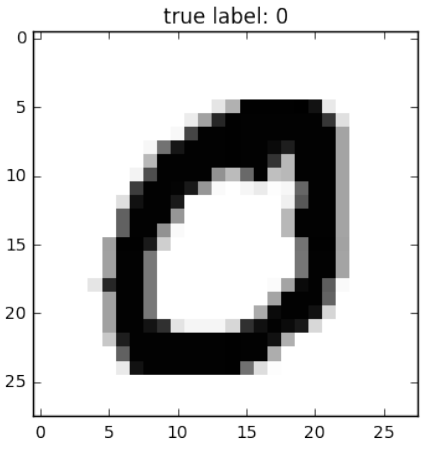
Example: Plain Vanilla NN



Let's assume that we have this type of neural network, and we train this network with the MNIST dataset. This dataset contains several thousand of handwritten numbers. In the end, the neural network can tell us which number is represented.

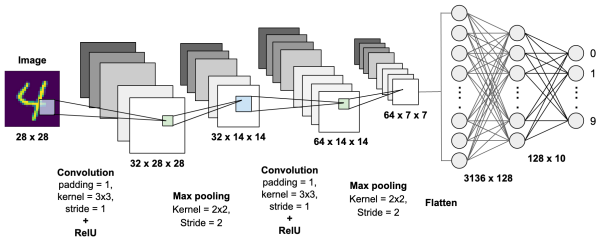
Input layer

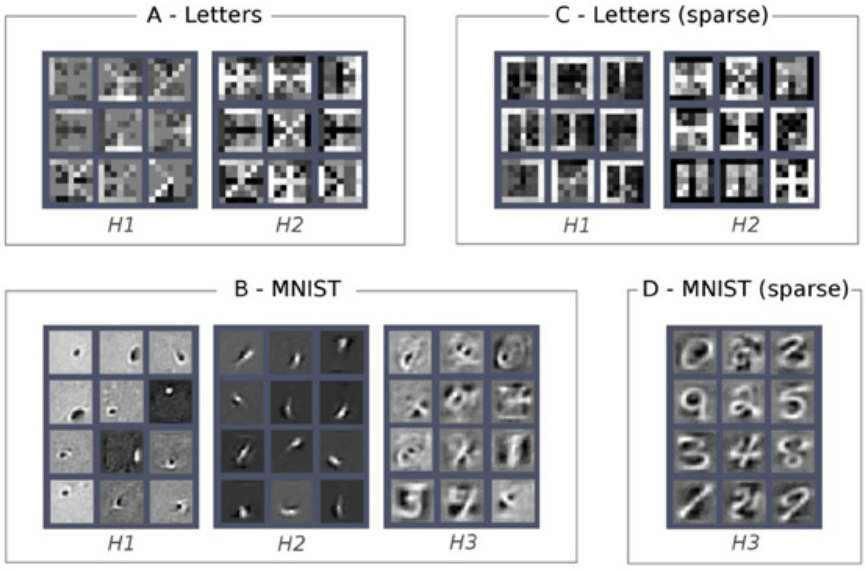
Let's say you have several examples of the number seven, one is depicted on the right. You could consider every pixel of this image as a neuron which contains a certain value. In this case, the value would tell us how bright the cell is. This is contained in the activation number of this neuron. Like that, you could take the value of a pixel and represent it as a neuron. This makes up the first layer of the Neural Network



Hidden Layers

The Hidden layers in this case are responsible for determine specific patterns with the aid of weights. By increasing and decreasing the value of weights, one is able to determine specific pattern. By increasing the amount of hidden layers, one is able to go more precise with these patterns (or obtain several matching pattern sequences).





Output Layer

The output layer consists of all the scalars of weight and activation for the specific number. The neurons which shines the brightest (the neuron which holds the highest number), contains the right output and therefore the solution.

The value of the output neurons can be summed by the following formula :

$$sigmoid(w_1a_1 + w_2a_2 + w_3a_3 + ... + w_na_n - 10)$$

The following formula in words would be :

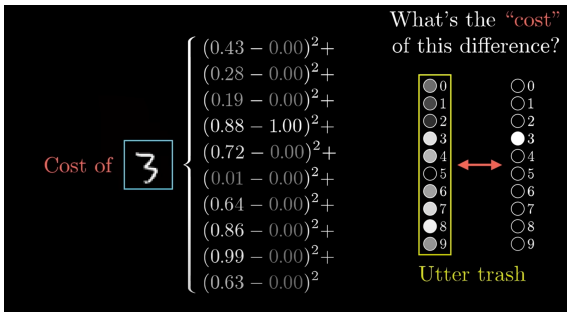
activation function(scalar product of all weight and sums - bias)

Learning process

A functioning neural network does not work perfectly without training. Indeed, the model will encounter many mistakes and therefore not the result one would expect. Basically, a neural network learns by its mistakes. The mistakes will lead to the adaption of weights (and therefore the sums), so that the weights are altered in the most convenient way.

The way this is done is with the aid of a so-called cost function. This is a way of telling the computer how extreme the mistake was.

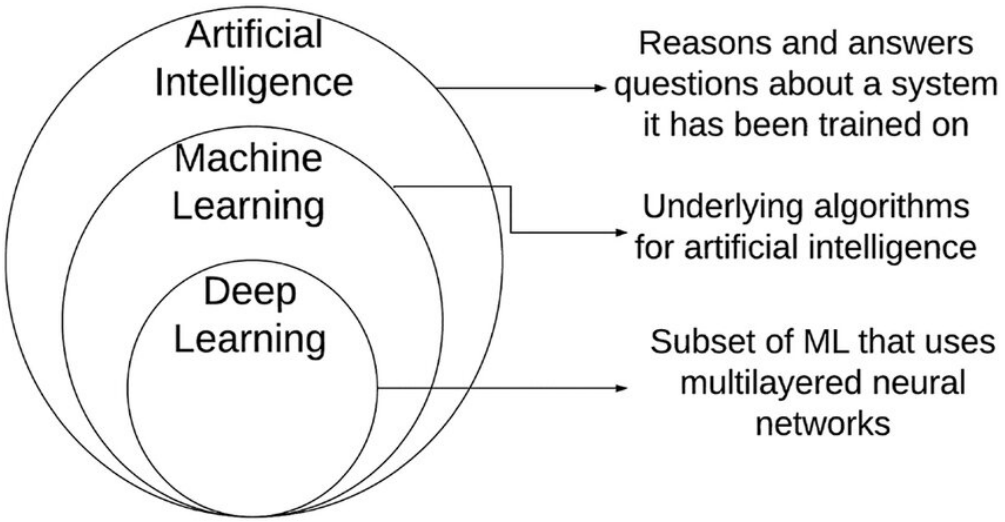
By adjusting and minimizing the intensiveness of the mistakes, we would get the proper weights and therefore the right output You add up the squares of the differences of the "mistake" output and the value you want them to have. If we go back to our MNIST example :



I we want the number 3, all other values of the output layers should have a value in the near of 0. A way how to minize these mistakes can be done with the aid of the gradient and the first derivative, which we will not go in detail.

Backpropagation

4 Venn Diagram



5 Appendix

https://ai-leaders.de/portfolio/wo-wird-ki-eingesetzt/#:~:text=Regierungen%20und%20Beh%C3%B6rden%20nutzen%20KI,zur%20Erzeugung%20von%20Kunstwerken%20einzusetzen.

https://www.youtube.com/watch?v=aircAruvnKk&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi

https://www.youtube.com/watch?v=4RixMPF4xis

https://www.youtube.com/@codingwithmagga

https://www.youtube.com/watch?v=q6kJ71tEYqM&t=386s

https://www.youtube.com/watch?v=tCApwsdijDk&list=PLNdSza1vBfZYh7gce5lctv1K1Xm3JjF-w

https://www.youtube.com/watch?v=W01tIRP_Rqs