\section{Neural Networks}

The following section is concerned with discussing neural networks as a means of investigating functional dependencies.\footnote{To aid with understanding the terminology used there is a glossary in the appendix section \ref{Glossary}.}\\

\subsection{General Introduction To Neural Networks}

%Definition

An artificial neural network (ANN) in the following called neural network, abbreviated to NN, is "a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs." \cite{NNPrimer} %\todo{cite: In "Neural Network Primer: Part I" by Maureen Caudill, AI Expert, Feb. 1989}\\

% Historical Context, First Introduction, Basic idea

First concepts of learning algorithms based on neural plasticity have been introduced in the late 1940s by D. O. Hebb \todo{citation needed}. In 1975 backpropagation became possible via an algorithm by Werbo\todo{citation needed}, this led to an increase in machine learning popularity. During the 1980s other methods like support vector machines and linear classifiers became the preferred and/or dominating machine learning approach. With the rise in computational power in recent years neural networks have gained back a lot of popularity.\\

The concept idea of neural networks is to replicate the ability of the human brain to learn and to adapt to new information. The structure and naming convention reflect this origin.\\

%Structue: Requirements, Deep networks

A neural network is made up of small processing units called neurons. These are grouped together into so called layers. Every network needs at least two layers, the input layer and the output layer. If a network has intermediary layers between input and output, they are called hidden layers. A network with at least two hidden layers is called a deep neural network (DNN). The amount of layers in a network is called the depth of the network. While the amount of neurons in a layer is called the layers width.

%As mentioned above the neurons are highly interconnected. The structure of connections determine the type of network. In the following we will discuss fully connected networks\footnote{Fully connected networks are also called dense networks.}. These feature a connection between each neuron of adjacent layers\footnote{The amount of connections between two layers in a dense network is therefore equal the the product of the width of adjacent layers.}

%Basic working principle, introduced non-linearity

In a typical NN information stored in neurons is transferred into the next layer by a weighted sum. The connected neuron of the following layer then applies a non-linear function, called activation function, to calculate it's final value. This process in repeated until the output layer is reached. The activation function as well as the amount and order\todo{reword "order"} of connections can vary in between layers. The system according to which a network is designed is called a network architecture. The most important architectures in the following work will be \textit{dense deep feed forward} and \textit{autoencoder}. To give insight into the basic working principle an example neural network is depicted and described in the following section \ref{NNExample}.\\

~\\

%\todo{General description of NN use-cases and strength compared to "old" machine learning methods such as support vector manchines.}

%Generally speaking neural networks are used to solve the equation $f(x) = y$ for $f()$. In other words it is used to make a fit to data points. There are many already established well known methods to do so. Therefore it is natural to ask what the advantages and disadvantages of neural networks compared to more traditional fitting methods are. To contrast neural networks we will consider support vector machines and analytical fits.

Neural networks are usually used in two ways, optimization or classification. Well known examples are handwriting recognition as classification and least mean squared error optimization\todo{think of better example}.

\subsection{Fully Connected}

\label{NNExample}

A fully connected or dense neural network like depicted in figure \ref{Img\_NNFully} is characterized by connecting every neuron from the previous to the following layer with a weight. In contrast to other networks this allows for a very high flexibility but also lacks the spatial context of data. This kind of network is especially well suited for data that is given in form of vectors or drawn from an arbitrary parameter space.%\todo{insert reference}

\begin{figure}

\includegraphics[width=\textwidth]{images/simpleNN.png}

\caption{Schematic structure of the most basic fully connected deep neural network. Indicated are the input (yellow), output (red) and hidden layers (blue and green). Each neuron outputs to all neurons in the following layer, but there are no interconnection between neurons of the same layer. Note that while the network has the minimum depth (2 hidden layers) to qualify for a deep neural network, the width could be smaller.}

\label{Img\_NNFully}

\end{figure}

%\todo{Funktionsweise}

The working principle is to form a weighted sum $\sum\_{i=1}^{N} w\_{i,j} \* x\_{i}$ over the values from neurons of the previous layer $x\_{i}$ weighted by the connecting weight $w\_{i,j}$. The weighted sum is then evaluated by the activation function $\sigma()$ such that the new value $x\_j = \sigma(\sum\_{i=1}^{N} w\_{i,j} \* x\_{i})$. Here i refers to the index of the neuron in the previous layer and j to the index of the neuron in the current layer.\\

Since forming a weighted sum is a linear operation the activation function must be non-linear to enable the network to learn non-linear behaviour.\todo{ref needed} The most common activation functions are the rectifier also called rectified linear unit (ReLU) and exponential linear unit (ELU) shown in figure \ref{ReLUELU}. Both are inspired by the asymmetrical behaviour of biological neural connections\todo{reference needed}.\\

\begin{figure}

\begin{subfigure}{.49\textwidth}

\centering

%\includegraphics[width=0.8\*\linewidth]{images/ReLU.png}

\missingfigure[figwidth=0.5\*\linewidth]{Picture depicting ReLU}

\subcaption{Rectified Linear Unit}

\label{ReLU}

\end{subfigure}

\begin{subfigure}{.49\textheight}

\centering

%\include{width=.8\linewidth}{images/ELU.png}

\missingfigure[figwidth=0.5\*\linewidth]{Picture depicting ELU}

\subcaption{Exponential Linear Unit}

\label{ELU}

\end{subfigure}

\caption{Example activation functions rectified linear unit (a) and exponential linear unit (b) used to introduce non-linearity into neural networks.}

\label{ReLUELU}

\end{figure}

%\todo{Describe training cycle}

%\todo{Mention Convolutional to contrast}

In contrast to fully connected networks convolutional networks apply a so called filter to the input. These filters consider the inputs of nearby neurons as well. They are usually used in image recognition and require data that has information stored in patterns, most commonly special patterns as in images. Fig. \ref{Img\_NNConv} depicts an exemplary convolutional network for image data and gives indication to it's working procedure.