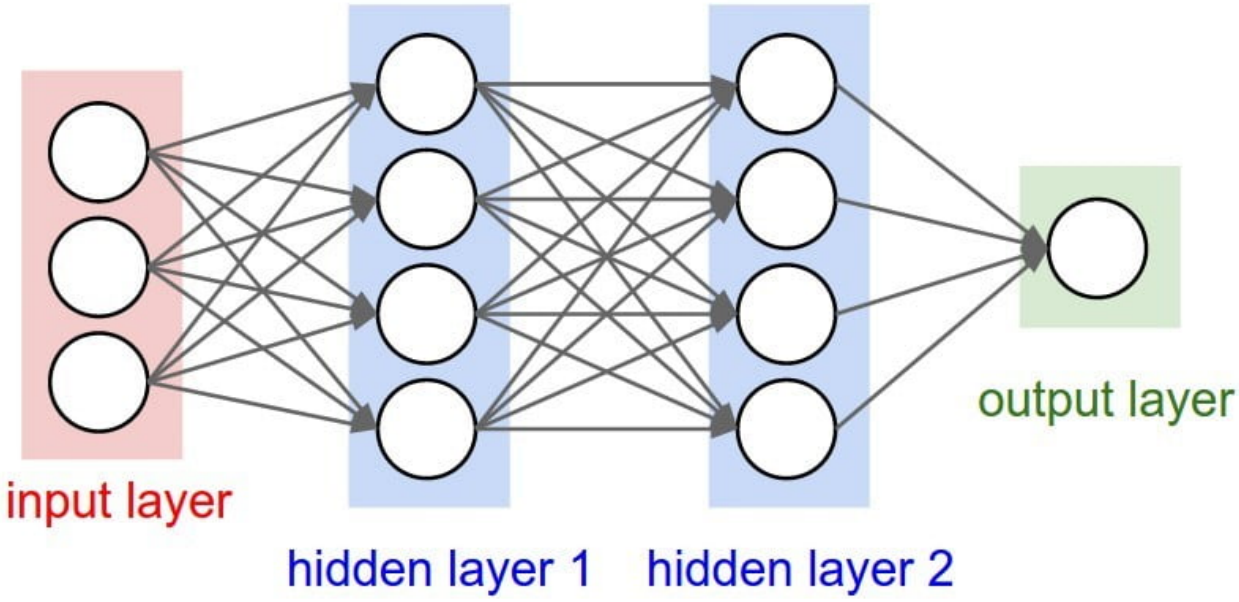
Mikias Berhanu | 2021280115
Assignment VI
Summary Artificial Neural Network, Overfitting, Underfitting and Cross Validation

Artificial Neural Networks

Artificial Neural Networks or ANN are designed to mimic the functioning of the human brain which allow computers to learn from data and make decisions without being intervened by humans. Their name is derived in reference to the human brain structure called neurons. ANNs are widely used in Machine Learning, Deep Learning and Artificial Intelligence. One thing to note here is that all the three mentioned fields are different in their implementation and what they stand for. ANNs are one of the algorithms used in Machine Learning and Deep learning.

Artificial Neural networks are composed of several layers of nodes or neurons where each node has its own purpose and contains weights and threshold value called bias which is later used for the decision making process. Usually ANNs start with input layers or input nodes and then middle layers which are commonly used for data processing and data sampling then the obvious output layer. Depending on the given task the output layer may give different results, for regression tasks the output will be continuous numeric value, for classification tasks it will be a probability between 0 and 1 for each class available in the data.



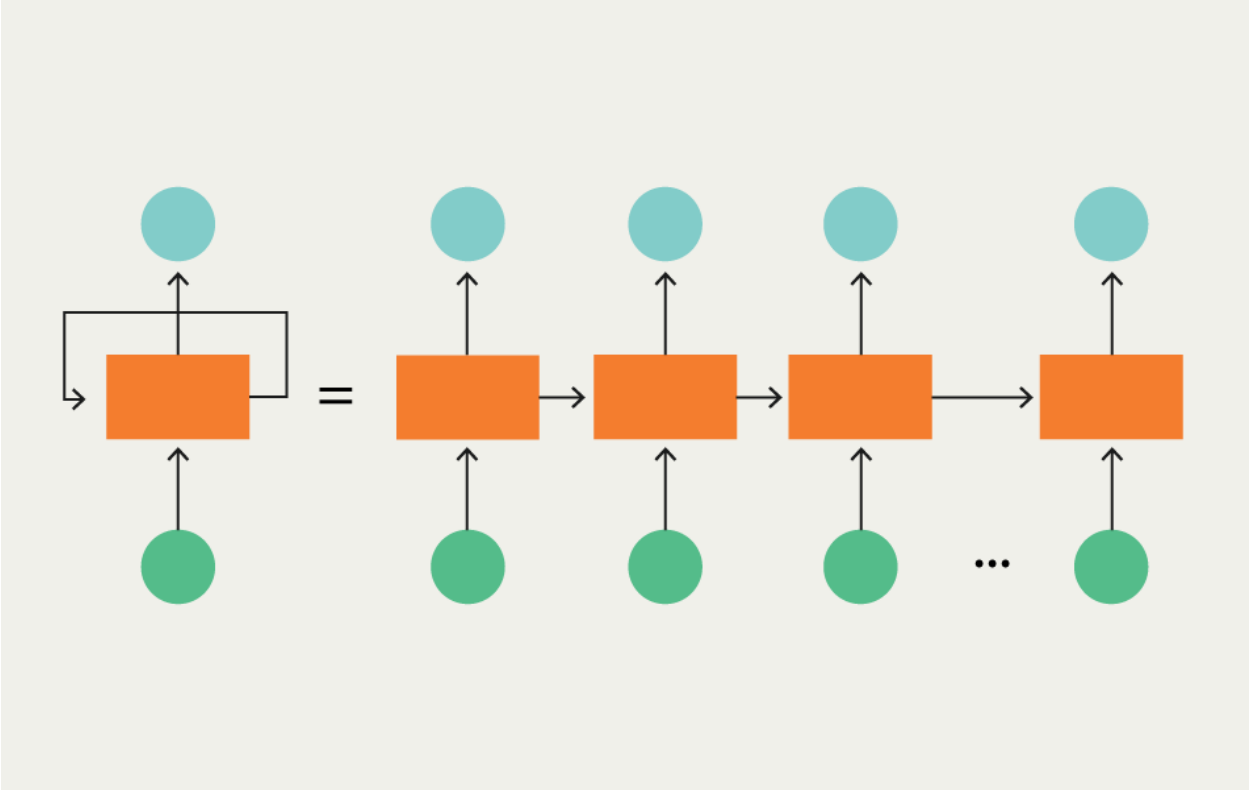
Artificial Neural Network Structure

In ANNs each input value has its own weight, weights are numeric values which describe how important a value is for the entire model. The higher the weight value the more significant the current value will be. Each input is multiplied with their respective weight value and summed up together which is basically matrix multiplication, in deep learning this kind of matrix operation is called Tensor operation. Once the inputs are multiplied and with the weights they are passed to an activation function which introduces non-linearity to the model. This will help the Neural Networks to generalize on data which is not linear by nature. THere are a number of activation functions but the most commonly used are Relu, Sigmoid …

A common thing in Deep Learning and Machine Learning is knowing the performance of our models which is very crucial for the entire process of model fine tuning. For ANNs we measure their performance using loss functions. And there are different loss functions depending on the task at hand. Mean Absolute Error(MAE) is used as a loss function for regression problems, Binary Cross Entropy is used as a loss function for binary classification tasks and Categorical Cross Entropy is used when we have multiple categories or classes to work with. The final goal of any deep learning model is to minimize the cost function, the model continuously adjusts the weights and bias to minimize the cost function during training. This process of adjusting the weights and bias to minimize the loss function is carried out using Gradient Descent Algorithm which is basically a derivation method for finding global minimums.

There are two different types of ANNs: Feed Forward Neural Networks, these Neural Networks are composed of an input layer, hidden layer and output layer. These Neural Networks include the popular ANN arthictectures like CNNs which are used in computer vision and in some cases for text processing. These networks are also called Multi-Layer Perceptron(MLP) which are composed of sigmoid neurons to handle the non-linearity of real world data. Recurrent Neural Networks: are another type of ANNs which have feedback loops. This type of ANNs are mainly used to handle data that comes in the form of time

series or time stamps like text, sound and voice processing, weather prediction, stock market prediction etc…

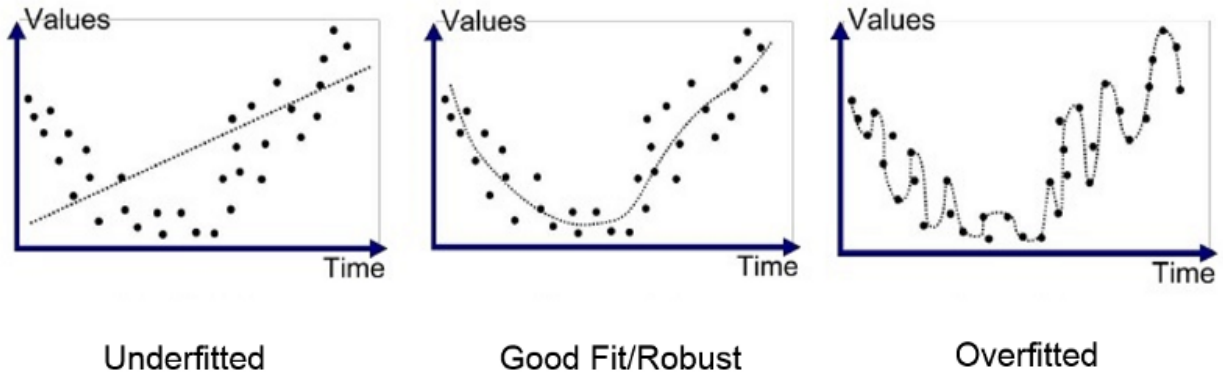


Recurrent Neural Network Structure

Overfitting and Underfitting

Overfitting is one of the common issues in Deep Learning and Machine Learning. This happens when the model performs really well on training data but fails to generalize on a testing dataset or new data that it hasn't seen before. Overfitting might happen for a number of reasons: noisy data, over tuning the model etc… This will affect the model's performance to generalize or properly predict on new and unseen data. Overfitting can be tackled in a number of techniques, in ANNs we can use two different techniques to overcome overfitting: Dropout layers, this layers will turn off or change the value of certain neurons to zero random which will be ignored by the learning algorithm, this will help the model to generalize better. Early Stopping, stopping the training phase before it reaches to over fitting, this will help to model to stop training and over fit.

Underfitting is the opposite of overfitting, here the model is not able to generalize both on training and testing set. Underfitting is easier to detect as long as we are using the proper performance metric. This can happen for a number of reasons as well: lack of training data, not using the right algorithm or model architecture for the problem at hand etc… This can be solved by using larger datasets for training, using different models and architectures etc…



Overfitting and Underfitting

K Fold Cross Validation

Cross validation is a way of evaluating the performance of machine learning models on a limited data sample. The data is sampled into K groups in which the model is later trained on. In order to perform K-Fold Cross Validation the data have to be shuffled randomly and splitted into K sub groups, then for each group we will have a training and testing set in which the model is trained on the training data set and then evaluated on the testing dataset. Once the model is trained and evaluated on all groups the results will be summarized.