AMITY UNIVERSITY
UTTAR PRADESH

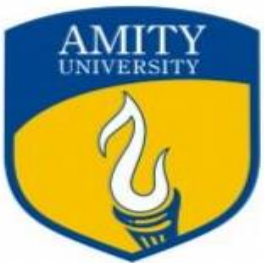
NTCC Term Paper

on

Deep Learning, Neural Network & its application.

Submitted to

Amity University Uttar Pradesh



In partial fulfilment of the requirements for the award of the degree

of

Bachelor of Technology

in

Computer Science and Engineering

by

Lakshya Tekwani

A2305219696

Under the guidance of

Dr. Stephan Thompson

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AMITY SCHOOL OF ENGINEERING AND TECHNOLOGY

AMITY UNIVERSITY UTTAR PRADESH

Declaration

I, Lakshya Tekwani, student of B.Tech (2-C.S.E.-11(Y)) hereby declare that the project titled “Deep Learning, Neural Network & its application.” which is submitted by me to Department of Computer Science and Engineering, Amity School of Engineering Technology, Amity University Uttar Pradesh, Noida, in partial fulfilment of requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

The Author attests that permission has been obtained for the use of any copyrighted material appearing in the Dissertation / Project report other than brief excerpts requiring only proper acknowledgement in scholarly writing and all such use is acknowledged.

Date: 15/05/2020

Lakshya Tekwani

A2305219696

2CSE-11(Y) (2019-23)

CERTIFICATE

This is to certify that Mr Lakshya Tekwani, student of B.Tech in Computer Science and Engineering has carried out work presented in the project of the Term paper “Deep Learning, Neural Network & its application.” as a part of First year program of Bachelor of Technology in Computer Science and Engineering from Amity University, Uttar Pradesh, Noida under my supervision.

_____

Dr. Stephan Thompson

Department of Computer Science and Engineering

ASET, Noida

ACKNOWLEDGEMENT

The satisfaction that accompanies that the successful completion of any task would be incomplete without the mention of people whose ceaseless cooperation made it possible, whose constant guidance and encouragement crown all efforts with success. I would like to thank **Prof. (Dr.) Abhay Bansal**, Head of Department-CSE, and Amity University for giving me the opportunity to undertake this project. I would like to thank my faculty guide **Dr. Stephan Thompson** who is the biggest driving force behind my successful completion of the project. He has been always there to solve any query of mine and also guided me in the right direction regarding the project. Without his help and inspiration, I would not have been able to complete the project. Also I would like to thank my batch mates who guided me, helped me and gave ideas and motivation at each step.

_____

Lakshya Tekwani

INDEX

Title : Deep Learning , Neural Network and Its Application

Abstract

Deep Learning algorithms imitates the working of the super powerful human brain in various aspects like processing data and creating patterns that allow decision making and therefore the structure of deep neural network is similar to that of the neuron network in the human body. Neural network comprises of multiple layers to gradually extract superior level features from the eggy input data. Artificial Network is the father of machine learning and in turn the Grandfather of Deep Learning, Deep Learning has networks which are competent of unsupervised learning from raw input i.e. unstructured data. Deep Learning possesses a superpower that is the ability to learn and therefore it can make better and efficient use of datasets for the purpose of feature extraction. Owing to its intelligence and ability to learn, Deep Learning becomes renowned aspect of Artificial Intelligence. Deep Learning is the secret ingredient that spices up many exciting and futuristic advancements and has made most of our wildest dreams and as well as the worst nightmares come true. Deep Learning applications may seem disillusioning to a normal human being, but to those with the privilege of knowledge and understanding of the machine learning world understands the dent that Deep Learning is making globally by producing machines with the ability to learn. Recently, Deep Learning has gained great progress in various fields like natural language processing(NLP), computer vision, Self-Driving Cars, Speech Recognition, Cybersecurity, Virtual Assistants, Healthcare, image & face recognition and many more.

Introduction

Introduction to Deep Learning and Neural Network

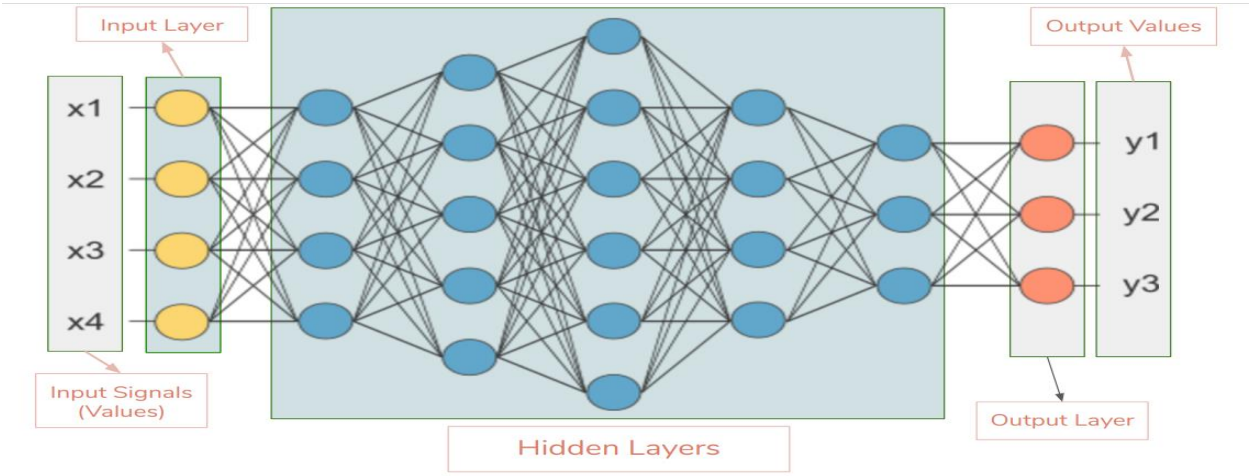
Artificial Intelligence is the all-inclusive concept that initially erupted, which was then accompanied by Machine Learning that flourished afterwards, and ultimately Deep Learning that is encouraging to elevate the advances of Artificial Intelligence to another level. Artificial Intelligence can be broadly elucidated as a mean to incorporate human intelligence into artificial machines. Machine completes task based on a set of predefined rules that solve various problems, such an “intelligent” behavior is referred as Artificial Intelligence.
Machine learning can be broadly elucidated as a mean to empower computing systems with the ability to learn. Its algorithms are such that they can learn how to make decisions.
Deep Learning is most convincing approach to Machine Learning, Deep Learning is a methodological ideology that focus at learning feature based on hierarchies i.e. a composition of lower level features results in the formation of features from higher levels of the hierarchy. Deep Learning allows networks to learn multiple features spontaneously at different level of abstraction to train complex function mapping of the input to the get required output by scrutinizing data, that too independent of being continuously programmed by the programmer.
Neural networks principle concepts are induced by the understanding of the anatomy of the human brain – all those interconnections between billions of neurons. At the same time contradicting a biological brain where any neuron can be connected to one another within a certain physical distance, millions of discrete layers, inter-connections between them and flow of data is well defined unlike the human neurons. Throughout the learning process synapses or signals are modified by the brain to enhance the behavior and responses. Both Machine learning and Deep Learning needs enormous amounts of data to function, the required data is stored by the billions of sensors utilized by the IOT(Internet Of Things) mechanism.

Machine Learning algorithmic design models enhances themselves progressively but the model still needs some manual guidance and time to time inputs to learn features. But if there is an error in the result predicted by machine learning model then it calls for a programmer to fix the algorithm manually but this is the point where Deep Learning takes the lead, as Deep Learning model does it by itself. Deep convolutional networks have completely changed the processing of video, speech, images, and audio, at the same time recurrent nets have enlightened the path of subsequent data that includes data like speech and text. Deep Learning is called so because of its enormous network of hidden layers as shown. All these layers are the reason behind the accuracy and the success of these deep neural networks.

Neural network comprises of a processing system with enormous interwoven layers and nodes that functions exactly as the neurons in a human brain. Every single neural network layer possesses a set of algorithms that venture to recognize relation within a set of data through a well-define process that imitate the working the human brain operation.
Deep Learning explores convoluted layout in huge data sets by the technique called backpropagation algorithm that controls how the initial parameters of a networks should be changed in order to minimize the error that

occurred when the signals were passed from layer to the other. Back propagation technique uses multi-dimensional and multi-utility feedforward network that is trained to reduce the error. The backpropagation technique adapts swiftly by computing and analyzing the synaptic updates by using feedback connections to deliver error in our output through signals. The signal is initially propagated forward so that the result can be generated with random initial parameters and then the error is transmitted backwards so that the random parameters can be chosen, in order to reduce the error - this was the revolutionizing introduction of backpropagation by the Geoffrey Hinton.

Deep Learning empowers the computational designs that are made up of several hidden processing layers that constitutes a deep network to learn interpretation of data with greater extent of precision. This has drastically upgraded the improvisation in various fields like speech recognition, computer vision, object recognition and various other fields such as drug manufacturing, healthcare facilities, automobile, security and genomics.



Literature review :

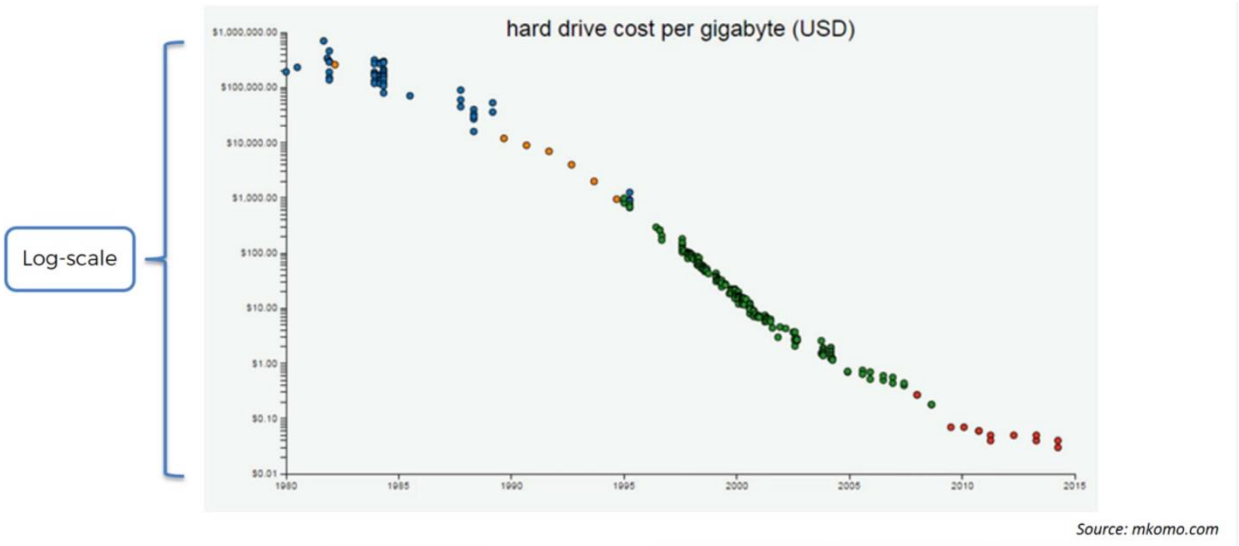
Artificial Machine learning and Deep Learning all of these have storytelling roots that penetrates as deep back as Greek antiquity. They have evolved from Science Fiction to Reality. John McCarthy was one of the few American computer scientists who started it all. McCarthy was one of the founders of the Artificial Intelligence discipline and he was the one who coined the term "Artificial Intelligence. From 1957 to 1974, AI flourished. Arthur Samuel, in 1959 coined the term Machine learning and then machine learning was not actually divergent from Artificial Intelligence. In the late 1970s and about the early 1980s, researchers had concentrated on adapting Artificial Intelligence for knowledge-based as well as logical approaches and not on algorithms. Apart from that they did not emphasized on neural network development, neural network research was abandoned by Artificial Intelligence and Computer Science researchers. This resulted in a chasmic difference between AI and ML researchers. Till then, Machine Learning was used to develop AI.

But after that the machine learning field, which included a large number of researchers and technicians, was re-introduced as a separate field, that too after years of struggle. There was a shift in priorities from programming for Artificial Intelligence to resolve and tackle practical and real-life problem in terms of providing services. During this era, the Machine learning industry focused on artificial neural networking and from then the development took pace.

Deep Learning along with neural networks have been around for quite a long time, neural network was discovered in the early 1940s. The roots of Deep Learning dates back to 1943, during that time W.H. Pitts & W.S. McCulloch created a computational design based on the working of neural system in the human brain. They used an amalgam of mathematics and layers of algorithms that they referred to as the threshold argumentation, which used to imitate the thinking mechanism of human mind. Steadily after, over the years Deep Learning evolved with two prominent highlights in its development. First was the formation of the ideology of the “Back-Propagation Model”, for that the credit goes to H. J. Kelley, in 1960 and the second, S. E. Dreyfus turned up with a straightforward version that was based on the chain rule of propagation in 1962.

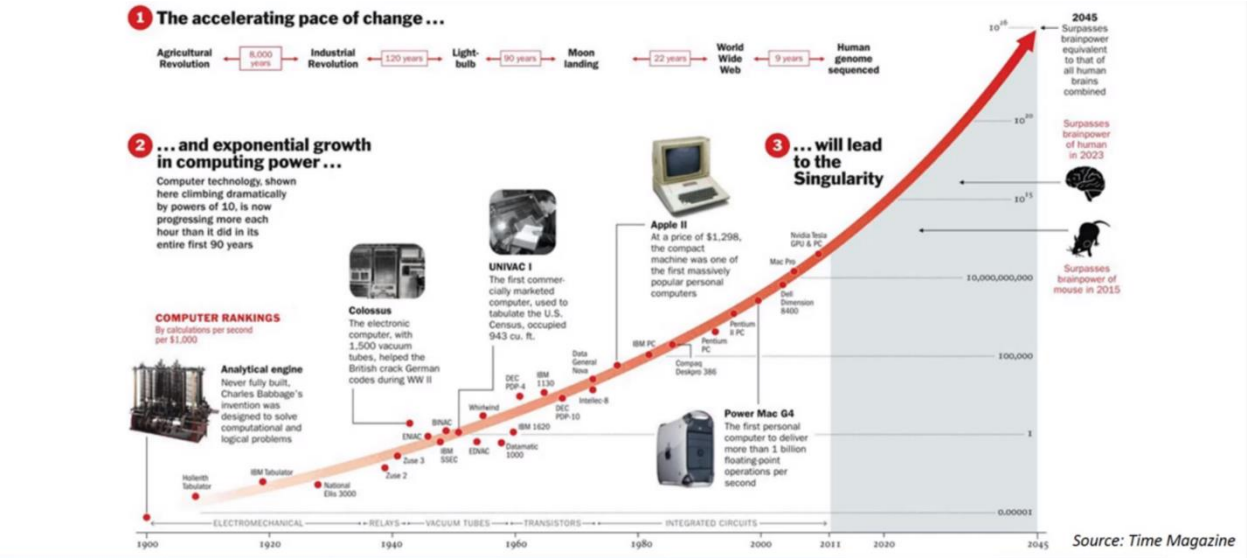
K. Fukushima was the first one who used Convolutional neural network, the neural network with multiple mapping functions and convolutional layers. K. Fukushima was among the ones who first developed an artificial neural network, named Neocognitron in the year 1979. Neocognitron uses a hierarchical and multi-layered pattern, which allowed the system to identify and detect visual patterns. The networks resemble a modern version along with a supplementing strategy of reoccurring activation by the various layers which robustness with time. But at that point of time they were not popular because of their requirements that were loads of data and processing power which was not available then and then the storage limits of memory had shown exponential growth and storage and memory costing have reduced significantly.

"Deep Learning" was coined by I. Aizenberg in the early 21st century. The current trend in data and memory is shown with the help of a graph of hard drive cost per gigabyte over the years.



The costing of memory storage gave reduced exponentially and the storage capacity have increased exponentially as it can be seen, therefore it was a hurdle for the development of neural network architecture because what it demands is only loads of data and processing power. Below is an image showing the exponential growth in storage. the first image is a 5 mega byte hard drive, the renting cost for that was nearly 2500 USD at that time, that's really high now the cost of hard drive is really low, now there are various online cloud storage platforms offering free storage.

Processing power , it was the second hurdle which inhabited the growth of Deep Learning at that time has also grown exponentially , in 2015 we passed the intelligence level of brain power of a mouse and by 2045 we may even achieve the processing power of human brain in a computational device .



Until the the implementation of ANN(artificial neural network) was only as a tool to help the system to recognize various types of patterns and replicate human intelligence by the end of 20th century .By the early 21st century, Yoshua Bengio, Geoffrey Hinton, and Yann LeCun were some of the very few enthusiasts who remained enact to their attitude towards Deep Learning. Although at that time their effort attempts to recharge the AI gang's interest towards the artificial neural networks at first were met with cynicism and went in vain but their ideology recently developed and advanced major technological advancements and their ideas like backpropagation, forward feed networking are now the trending in the field. Geoffrey E. Hinton's Publication titled "How neural networks learn from experience" gives a proper algorithm of how back-propagation algorithm works and that has revolutionized the whole concept of deep learning i.e. how a machine can learn from its mistakes. Yann LeCun , Yoshua Bengio & Geoffrey Hinton in their publication titled "Deep learning" stated "Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction " , deep neural network are composed of numerous hidden layers and forms a synaptic web between them , information passes from one neuron to other with some weights attached to them and this results in giving a precise output. Various other algorithms were proposed, but none of them was as effective as back propagation model which was designed by Geoffrey E. Hinton, due to his contribution, he is also referred as the father of Deep learning. Jiquan Ngiam, Aditya Khosla, Mingyu Kim , Juhan Nam, Honglak Lee, Andrew Y. Ng in their publication titled "Multimodal Deep Learning" presented a way for cross modality feature learning ,which tells that learning output is better when more features are used for one modalities , more the features ,more will be the learning rate and optimization .

Methodology

Neural network algorithm

Neural network algorithm is most powerful and robustly used algorithms in the subdivision of Machine Learning referred to as Deep Learning. Apparently, Neural Networks appear like a room full of darkness, the input layer receives the data and transmits the data to the mysterious “hidden layers” and later performing a piece of magic we may get the output provided by the end or the so-called output layer. Grasping the magic these numerous hidden layers are performing is the main essence in the learning of the neural network.

Neural network performs a series of computing algorithm that comprises of a network of straightforward yet highly integrated nodes which are assembled as layers which processes data through a process that replicates the functioning of the human brain cells, neurons work. The brain has approximately 100 billion neurons so there can be any number of hidden layers.

The data is received by the input layer, which in turn assigns the data to input neuron layers, each component of data into different neuron ,through these neurons and synaptic signals the data is them transmitted to the enormous network of the magical hidden layers , they are called ‘hidden’ due to the fact that they do not compromise the input or output layer. Hidden layers is the mysterious location where all the computation and processing actually happens through a network characterized by parameters :which are the Weights (W) and the Biases (b) , the input layer receives the data , then it transmits the data to the hidden layer neuron , these calculate a total weighted sum of the input by summing up the products of the argument with their corresponding weight and then finally adding the corresponding bias and this gives a input value for a pre-defined activation function ,it processes the data and decides whether it should be discharged or activated. Subsequently, the neuron transmits the data downstream to other connected neurons by a process named ‘forward pass’. After all this, the data is transmitted to the output layer which assigns separate neuron for each desired possible output.

Perceptron is a neural network that consists of a single layer, the perceptron consists of 4 parts.

- The one and only input layer
- Its corresponding weights(W) and biases (B)
- The net or the total weighted sum of the input arguments
- The corresponding activation function.

The process is as follows:

a. All the inputs values **(x_i)** are multiplied with their weights **(w_i)** call them **k_i**

b. Add all the **k_i** values and call them Weighted Sum(**W**).

c. Apply correct Activation Function to that weighted. Determine whether the value after the application of the activation function is greater than a particular threshold reference, where **the negative of (-)** the threshold is equal to bias, and returns one (1) or zero (0) as the output.

Notes: (**w_i**). **(x_i)** is the scalar product of **(w_i)**, a vector whose component are weights, and **(x_i)**, a vector comprising of the input values.

We can vary values associated with the randomized parameters like the weights(W) and the biases(B) to obtain perfect patterns of executive models. More weight (higher values can be

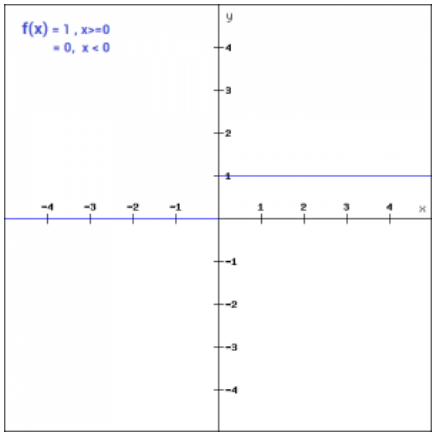
associated to them) can be assigned to those inputs in such a way if they give outputs greater than zero, it will favor output which we want or the correct output.

The value of the bias is the measure of the difficulty or the ease with which the output value gives the value 1, we change the value of the biases as per our observations, we can assign the values of bias on the basis of a particular threshold limit. If the bias is big and greater than zero then it will make our output tend towards 1; however, a bias with value less than zero of will makes out results tends towards 0 or false. In general we actually create a deep web of layers or a networks that comprises of a large number of perceptrons where each perceptron layer receives the input which is indeed the output of the preceding perceptron and applies the weight values to the output and make a complex decision.

The most common problem related to a network made up of only perceptrons is that a small change in input can actually change the output from zero to one or vice versa. But for our purpose we need to gradually increase the or decrease the input and the results should also gradually increase or decrease. Therefore, a different and more modern type of neurons come into play each associated with a different activation function:-

Binary Step Function

This function gives the result based on the observation that if the input value is greater than a particular threshold value , then the neuron gives the output as true (1), or else it gives false (0), i.e. its output is not passed to the following and it is assigned a value zero. This is basic activation function which we have discussed. The problem associated with this function is that it is not continuous.



2. Linear Function

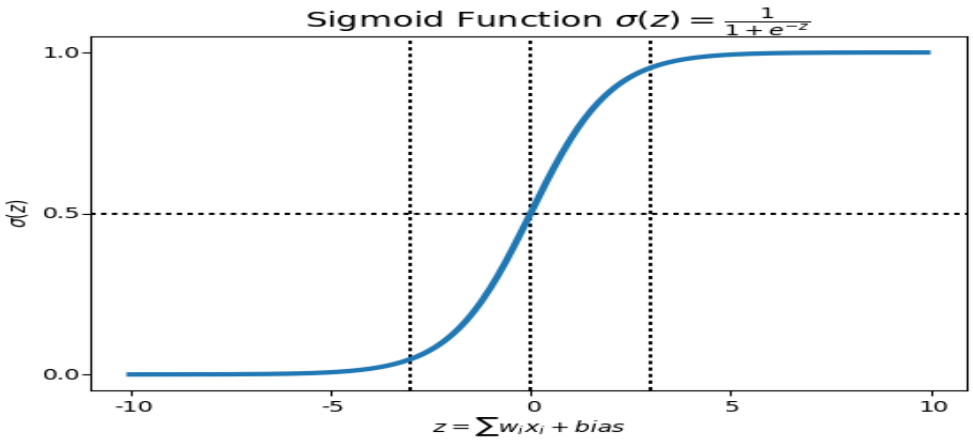
The problem associated with the binary step function i.e.; the gradient or the partial derivate with respect to x becomes zero because there is no x component in the above discussed function. Instead a liner function is used. Which is defined as

f(x)= az +b (a is any constant and b is bias)

3. Sigmoid function

Sigmoid function is far better than binary step function or linear function due to the fact it can give continuous value between 0 and 1. When z (weighted sum) is positive and large the function gives the maximum permissible value i.e. 1; on the other hand if z(weighted sum) is negative and small (mathematically), the function gives its lowest permissible value of 0.

The sigmoid function is better because it can give a value between a range of 0 to 1 instead of two particular value 0 or 1 this is the fact is superior to the above two functions. A small change in input gradually changes our output; this was the ideal behavior we were looking for a perceptron.



The above graph shows the effectiveness of sigmoid function.

The ability to learn every time when the mechanism predicts an output, is the core strength of the machine learning algorithm. Learning in this context refers that the weights and biases that corelates neurons tends to become more accurate and precise, the main aim of selection of the weights and biases is to approach the real value y(x) for all the training inputs. This refers that the output $\bar{y}(x)$ which the neural network gives is very close to the real value y(x) .

Our objective is to reduce the error , to achieve precision therefore we need to find the difference between our predicted result and actual result there fore we need a particular function for that ,i.e. a cost function , a cost function is nothing but the error in predicting the correct output in out network ; in other terms, it is the difference between the actual value and the predicted output. In neural networks, the most commonly used one is the quadratic cost function, also called mean squared error but there are various types of cost functions, which are enlisted below:

- Mean Squared Error Cost Function (MSE)

- Cross Entropy Cost Function
- Kullback-Leibler (KL) Divergence

The best one among the above function is the mean squared error function over the functions due to the fact that in neural networks small changes in weights and biases do not produces any change in the number of correct outputs; so by using a Mean Squared Error function which is quadratic function helps to create big differences which have more effect on the cost function instead of small ones which will help figuring out how to modify the parameters (weights and biases).

The value of cost function minimizes when output value $\bar{y}$ is closer to the real value y, for all training inputs. The main aim of our network is to minimize the cost function and our output value should approach to the actual value, the main tool to achieve this goal is an algorithm called Gradient Descent.

We know that a function can have global maximum or minimum, that is the point, at which the function attains the maximum or minimum value. One way to obtained that point is through the derivative method . However, it seems easy only when we need to calculate a function with two variables but in the case of neural network, they include a lot of variables which make this computation quite impossible to make.

For example let's consider we 25 weights and lets assume only 1000 possible values that can be associated to each weight .

Total combination= 1000 X 1000 X 1000 X……X1000= $1000^{25} = 10^{75}$

because each weight can have 1000 possible values.

According to TOP500 project ranking in November 2019 , world's fastest computer is **Summit or OLCF-4**. It is capable to perform at a speed of 200 peta FLOPS(floating point operations per second). That is equal to 200 X 10^{15} FLOPS . lets assume 1 FLOP for 1 weight possibility , its actually more than that .

So the time required by **Summit or OLCF-4** to find the correct set of weigths =

$$\frac{\text{total number of operation}}{\text{Number of operation per second}}$$

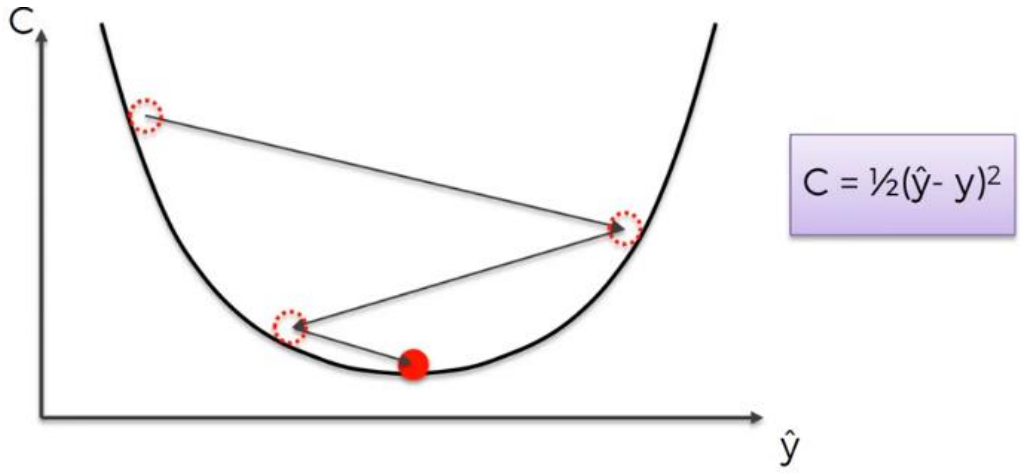
$= 10^{75} / 200 \text{ X } 10^{15} = 5 \text{ X } 10^{57}$ seconds

`$= 1.58 \text{ x} 10^{50}$ years

Age of earth = 1.54 billion years = 1.54 $X10^{9}$

So the time required by the words fastest computer is even more than the age of the earth , so it's impossible to try all the combinations therefore we need a more practical approach . our neural network assumptions are very less unlike the neural networks used every day, therefore they have more and more hidden layers

let's take a quick look at the graph of a random function,



This function has a global minimum. We can find the partial derivative of the function to calculate where the minimum is located or we could take another approach.

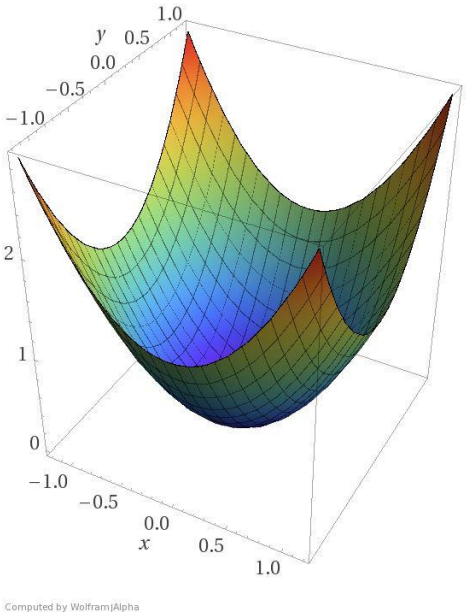
When we initialize our weights randomly, we are at point A in the cost function. firstly, we check out all possible directions in the x-y plane, and to find the direction which results in steepest decline in the value of the cost function, continue to move int that direction. the direction of the gradient is just opposite to this direction. The gradient or the so called the higher dimensional cousin of derivative, gives us the direction with the steepest ascent. Gradient of a function is Vector of partial derivatives.

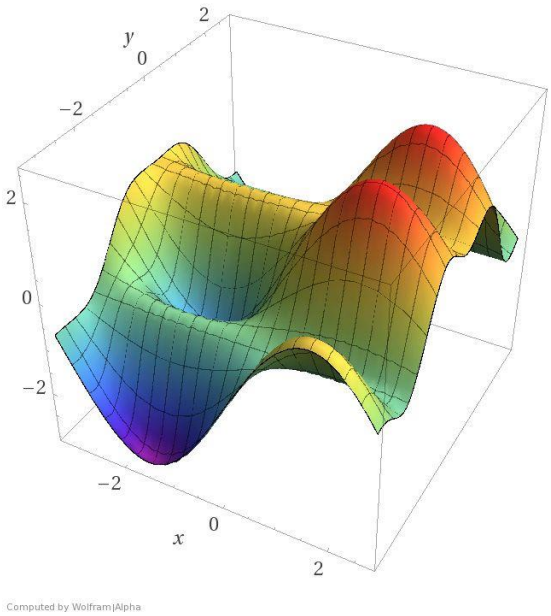
Once we have discovered the direction in which we want to move, then we need to select the size of the step to be taken .The size each of this step is called the learning rate. We should ensure that the chosen value of the learning rate is neither too big nor too small to ensure we can get down to the minima.

If we go too fast, we might overstep the minima, and keep roaming along the ridges of the "valley" without reaching the minima ever.Going too slow, will result training turning out to be too long to be feasible or practically possible. very slow learning rates make the algorithm more prone to get stuck at the minima.

Once we have our gradient and the learning rate, we take a step, and recompute the gradient at whatever position we end up at, and repeat the process till we reach the minima as shown in the graph.

As the steepest ascent is in the opposite direction of gradient, its magnitude tells us how steep the ascent/descent is. Therefore, at the minima, where the contour is almost flat, we expect the gradient to tend towards zero, it's precisely zero for the point of minima. But it’s really difficult to find the minima as shown in the given diagrams.





The method of Gradient Descent works by determining the Gradient ∇C continuously of that particular function and at the same time updating the weights and biases, our aim to find the correct values that minimize the error and hence reducing the cost of function value. And through this process the neural network learns.

Sometimes, calculating the gradient can be very complex. Therefore, different types of gradient algorithm are used some of them are listed below:

- **Batch Gradient Descent**
- **Stochastic Gradient Descent**
- **Mini Batch Gradient Descent**

The main objective of this algorithm is to find out the partial derivatives of the cost function with respect to any weights(w) and biases(b) ,to reduce the cost function , and for that reason we need we calculate the error vectors starting from the final layer and then, propagating backwards to update the weights and biases and this is called back propagation. The reason behind this backflow mechanism or why we need to go back is that the cost is a function of the output of our network and hence to change correct the error we need to start our calculations from the final layer. Backpropagation a central mechanism through which neural networks learn itself. Messenger tells the network whether the net made a mistake or not, when it made a prediction. A neural network propagates the input data and its parameters through a feed forward network towards the moment of decision, and then backpropagates the error about the output, in reverse through the network and then the network alters the parameters (weights and biases).

An overview of the neural network algorithm is depicted below:

STEP 1: Assigning the values to the weights and biases near to 0 (but not 0).

STEP 2: pass the input data to the input layer and one feature to each input node.

STEP 3: Forward-Propagation: propagating from left to right, the neurons are activated by the activation function in such a way that the impact of each neuron activation is limited by the weights until the threshold value or bias value is reached. activation function is propagated until getting the predicted result $\bar{y}$.

STEP 4: Compare the predicted output $\bar{y}$ to the actual result y. Measure the generated error.

STEP 5: Back-Propagation: propagating from right to left, the error is back-propagated. Weights are back propagated according to how much impact they affect for the error. Then the weights are updated according to learning rates.

STEP 6: Repeat the Steps from 1 to 5 and update the weights each time after each observation (Stochastic Gradient Descent method)

Or else

Repeat the Steps from 1 to 5 but update the weights only once after a batch of observations (Gradient Descent).

STEP 7: When the whole training data set is passed through the artificial neural network, the desired output will be obtained with minimum error parameters are reduced for future observations.

Results

As we have discussed the applications and algorithms of Deep Learning there are various problems faced by the Deep Learning technique, the first and the foremost problem is locating large training datasets. Although, this is not an issue for consumer applications where vast quantities of data are readily accessible, ample amounts of training data are rarely available.

Secondly the problem is that it can take days on large clusters of CPUs and GPUs graphics processing units using them to train deep-learning networks, once the data sets are in hand. Emerging strategies such as transfer learning and adversarial generative networks show some promise in overcoming this obstacle.

Third, one of the reasons why Deep Learning works this well is the plenty of interlinked neurons or free parameters that allow slight nuances and data variations to be captured. This implies that hyperparameters and those parameters of which we values need to be set before training, are more difficult to define. The method is one that is more art than science. There is also the risk of overfitting data, particularly when the set of parameters surpass the considerable number of independent observations.

Fourthly, it is hard to comprehend why Deep Learning networks reaches at insights due to the increasing amount of layers, nodes and connections. While it is not so significant in applications such as tagging images on various social media platforms, recognizing the judgment-making process in objective oriented applications like clinical decision- making is very important. Researchers along with some US military personnel are working on developing "explainable AI" to resolve this problem.

Lastly, deep-learning networks are extremely vulnerable to the butterfly effect i.e. any small variations in input data will result in significantly different outcomes, making them unstable. This instability also opens new surfaces for hackers to attack. Researchers have shown in what are known as adversarial attacks that, simply by introducing an insignificant amount of noise. Deep Learning networks can be tricked into arriving at totally incorrect conclusions – all without even accessing the framework.

But apart from that Deep Learning (DL) gained overnight success when a robot player won against a human player in the famed game of AlphaGo. Deep Learning instructions and learning methods have been widely acknowledged for “humanizing” machines. Due to the rapid growth of machine many of the advanced automation capabilities are now found in enterprise AI platforms.

Deep Learning (DL) has a great scope in near-future and the biggest reason behind this is that it does not require any type of on hand engineering. Deep Learning doesn’t require an instructor giving it the features after extracting it from the data as it extracts the features from data itself. And in this way our biggest problem of

feature engineering is solved. Also, since features acquire skills from the model itself, it has a greater probability of producing a more generalized model than the feature engineered models.

These reasons alone are acceptable to prefer DL over other technologies. With the latest advancement, we now have futuristic outcomes in various fields including probability and statistics , progenitors, computational intelligence, data mining, data science ,Natural Language Processing, Language Translation, Speech Recognition, Image Recognition and Classification, Language generation among others, Artificial Intelligence has grown more and more stronger in past 6-8 years with Deep Learning than it has developed in past 20 years before that. The tremendous growth in Deep Learning research and industry applications reveals its "omnipresent" nature in every facet of Artificial Intelligence — be it machine learning or security or computer vision applications. In near future with increasing research opportunities, unsupervised learning methods will be able to deliver models that will closely mimic human behavior and intelligence. With advancement in technology, we can make CPUs and GPUs that are cheaper and faster, which can enable the production of bigger, more efficient algorithms and design neural nets capable of processing more data and that to at a great speed, so it may learn to recognize patterns easily.

Due to the apparent conflict between consumer data protection laws and needs of high volumes of consumer data for the Deep Learning evolution there are various amendments in favor of Deep Learning. Better generative and responsive models will be created along with algorithms that can instinctively learn how to generate images, speech, text that humans can't differentiate from the real thing .Google's possession of DeepMind Technologies holds a promising future for global marketers in this field. Hundreds of industries could effectively use neural nets to function more efficiently, develop new products, target new audiences or improve consumer safety. We already communicate with smart speakers, which will become much smarter. Obsolete will replace keyboards and they will take place in museums beside typewriters. This will makes the benefits of Deep Learning available to all of us.
In future neural network might allow:

- robots with vision, sense and ability to predict the world around them.
- Smart virtual opponents in computer games
- World of self-driving cars
- automatic transformation of handwritten documents into text documents.
- human genome to aid in the understanding of the data compiled by the Human Genome Project.
- neural network usage for self-diagnosis of various health problems like tumors , cancers and various other types of illnesses .

Conclusion

We are on the verge of creating the most advanced machines in the history of mankind. The Deep Learning is a top-notch topic in the field of Science and Technology. Machines are getting smarter day by day because now they can learn from their own mistakes, in a way that is very similar to humans. Artificial Intelligence, machine learning and especially Deep Learning can change our lives completely with their super intelligence every task would be very easy to perform, and a major chain of technology will follow after we develop them efficiently. Deep Learning has its benefits as well as risks with which our researchers, scientist and engineers are well versed with. Artificial Intelligence will be improving each and every day. May it be Microsoft's Cortona or Apple's SIRI or Amazon's Alexa all have made our lives easier, even they will also get advance day by day as various companies invests billions of dollars in such an advancement. Deep Learning may not be the only savior of AI solutions in spite of being globally popular. Developers may soon find themselves outpaced if Deep Learning technology progresses in the current pace and hence will be forced to take intensive training. Enormous data is emanating from sensors, the network turn data into information, information into knowledge, knowledge into understanding and finally if we are fortunate then knowledge into wisdom. Deep Learning networks are connection between digital computers and the real world; this permits us to communicate with computers on our own terms

.

References

1. https://www.researchgate.net/publication/318447392_DEEP_LEARNING_A_REVIEW
2. https://machinelearningmastery.com/what-is-deep-learning/
3. https://ieeexplore.ieee.org/document/7804882
4. https://towardsdatascience.com/introducing-deep-learning-and-neural-networks-deep-learning-for-rookies-1-bd68f9cf5883
5. https://en.wikipedia.org/wiki/Deep_learning
6. https://towardsdatascience.com/clearing-the-confusion-ai-vs-machine-learning-vs-deep-learning-differences-fce69b21d5eb
7. https://www.geeksforgeeks.org/artificial-intelligence-vs-machine-learning-vs-deep-learning/
8. https://www.nature.com/articles/nature14539'
9. LeCun, Y., Bengio, Y. & Hinton, G. Deep Learning. Nature 521, 436–444 (2015).
10. https://doi.org/10.1038/nature14539
11. https://towardsdatascience.com/deep-learning-algorithms-the-complete-guide-ce020bd47ecc
12. https://towardsdatascience.com/deep-learning-algorithms-the-complete-guide-ce020bd47ecc
13. https://towardsdatascience.com/understanding-neural-networks-what-how-and-why-18ec703ebd31
14. https://mc.ai/32-advantages-and-disadvantages-of-deep-learning/
15. https://www.eetimes.com/5-challenges-for-deep-learning/
16. http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/
17. https://becominghuman.ai/what-is-deep-learning-2dd6192c1de7
18. https://icml.cc/2011/papers/399_icmlpaper.pdf (Multimodal Deep Learning)
19. http://www.cs.toronto.edu/~hinton/absps/NatureDeepReview.pdf (Deep learning)
20. http://www.cs.toronto.edu/~hinton/absps/sciam92.pdf (how neural network learn from experience)