Most of us know Deep Learning to be a 21st Century invention, but believe it or not, it has been around since the 1940s.

The reason most of us are unaware about Deep Learning advancements/researches of 20th century is because the approaches used back then, were relatively unpopular due to their various shortcomings and the fact that it has had a couple of re-brandings since then.

New original research in any field requires an understanding of the history, evolution and major breakthroughs that led to the popularisation of said field. Deep Learning is no exception.

A broader look at the history of Deep Learning reveals 3 major waves of advancements:

Cybernetics — During 1940–1960

Connectionism — During 1980–1990

Deep Learning — Since 2006

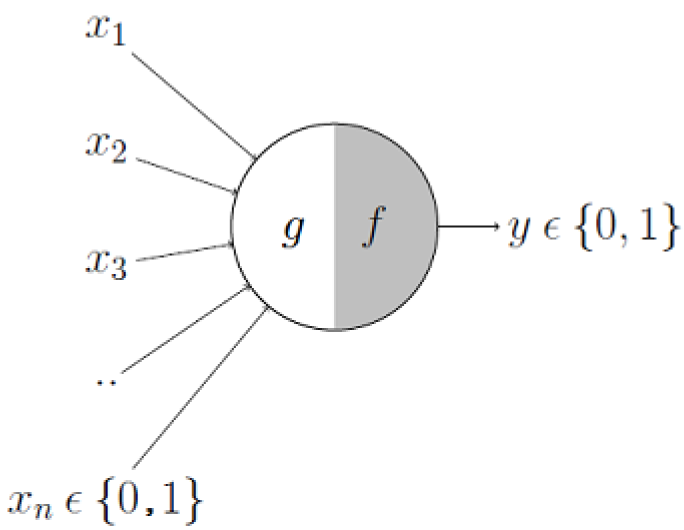
Researches done during the first 2 waves were unpopular due to the critics of their shortcomings, however, there is no doubt that it has helped advance the field to where it is today and some of the algorithms developed during those times are used widely till today in various machine learning and deep learning models.

Let’s explore the 3 waves in some detail to get some perspective.

Cybernetics

Is the earliest predecessor of modern Deep Learning, based on the idea of ***biological learning*** — how human brains learn. Advancements done in the name of cybernetics were based on the goal to replicate the working of a human/animal brain in a simpler computational model, which would help build systems that would start learning like actual brains and provide conclusions given some input. Further research under this mindset continues separately till now under Computational Neuroscience.

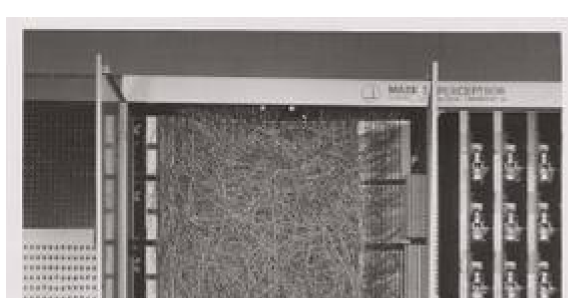
Cybernetics was kicked off by the development of the **McCulloch-Pitts Neuron.** It was an attempt to mimic the biological neuron. It was based on a linear model that would take various inputs [X1, X2 .... Xn], for each input the model had some weights [W1, W2 ... Wn] and the output f(x,w) = X1W1 + X2W2 + ... + XnWn. This model could only output True/False based on the inputs and weights.

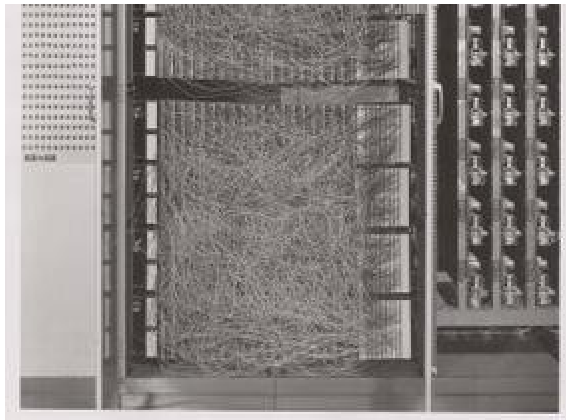


McCulloch-Pitts Model from Towards Data Science

The weights needed to be set properly, and were input by a human manually.

Later in the 1950s, **Perceptron** was developed by Frank Rosenblatt, an American psychologist, that would learn the weights automatically. Perceptron was initially designed to be an electrical machine rather than a program/software. Frank built Perceptron for image recognition, it contained photocells(receivers) connected to multiple neurons which would classify the inputs captured by the photocells.

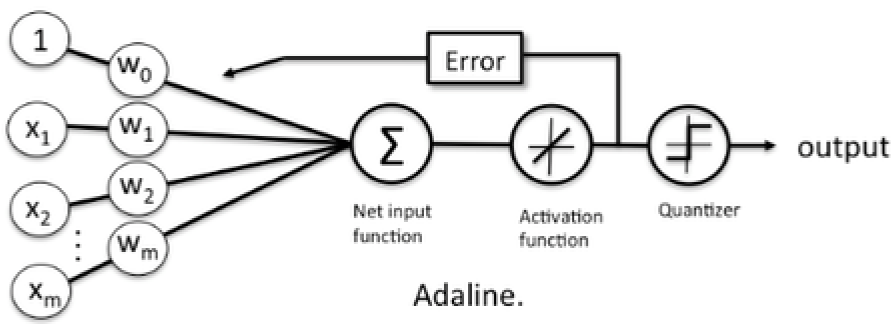




Perceptron from wikipedia

Though Perceptron was a remarkable machine for that time, it made bold claims which couldn't be fulfilled at the time.

ADALINE —Developed by Bernard Widrow, known as adaptive linear element, which was developed around the same time as Perceptron, could also adapt to the weights based on the weighted sum of the inputs during the learning phase.



ADALINE from mlxtend

The learning function for ADALINE is similar to stochastic gradient descent used in Linear Regression today.

These Linear models had various limitations and the critics who saw these limitations caused major dip in their popularity and stagnate the research for a while. One major limitation was that these linear models could not train for XOR functions.

Because these models were inspired by neuro-scientific research, the dip in their popularity also inspired exploration of models apart from neuroscientific basis.

Connectionism

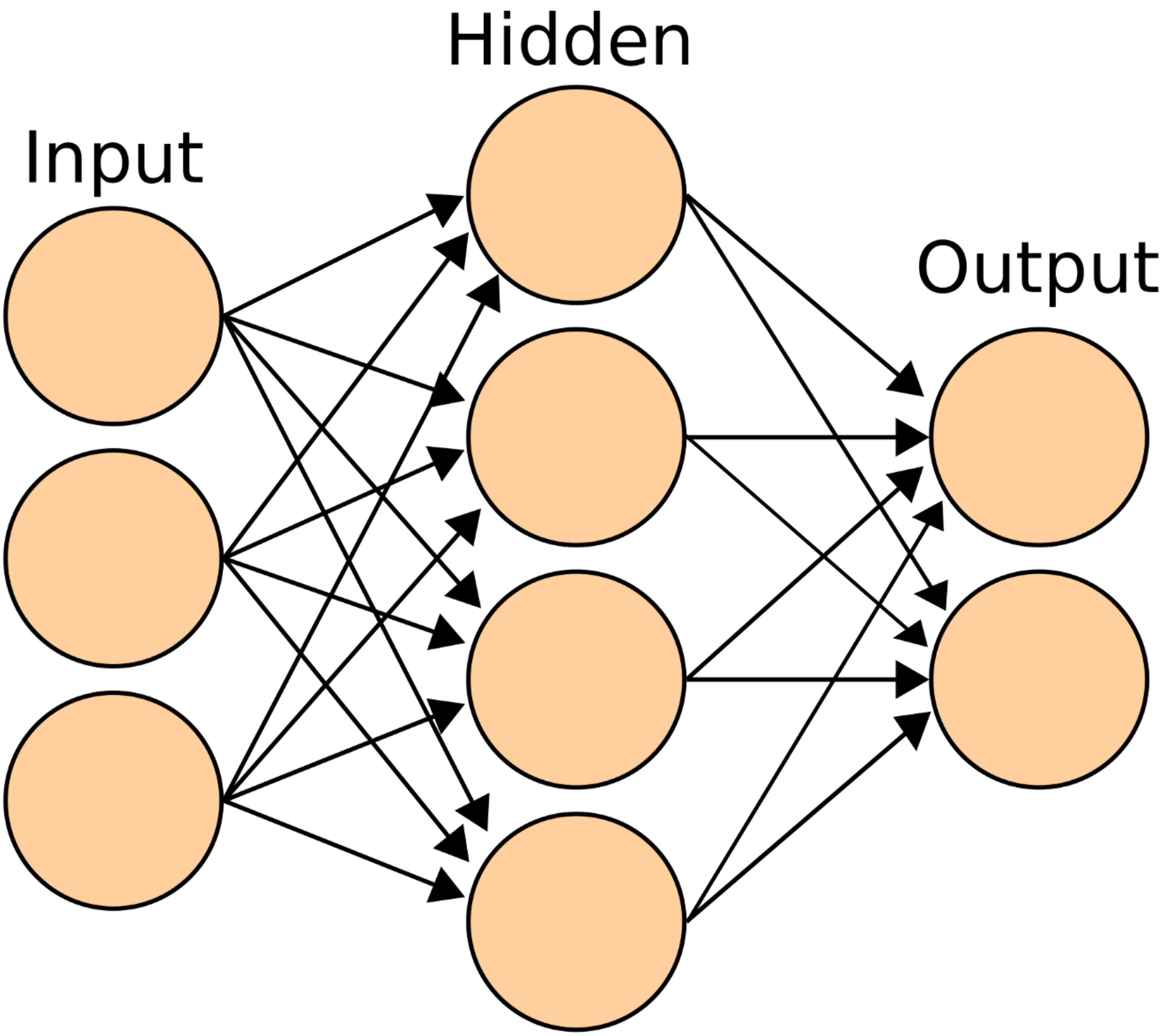
Or ***Parallel Distributed Processing*** became popular in 1980s. This approach was inspired by cognitive sciences. Connectionism showed promise compared to various symbolic reasoning approaches scientists were exploring in the 1980s known as Classicists.

Even though if we look at the brain in a more abstract view, Symbolic Reasoning approach fits well but are hard to implement explicitly using classical programmatic models. So practical connectionists viewed their work as leveraging Neural Nets to achieve similar effect as Symbolic Reasoning.

But Radical Connectionist simply discarded the idea of Symbolic Reasoning stating that it could not explain various complex features of our brain anyway and was an incorrect perception of the human brain.

Concept of Artificial Neural Network(ANNs) was introduced during this wave. The main idea behind ANNs was to develop a network of individual units that can be programmed to achieve intelligent behaviour. This was the first time the concept of *hidden layers* was introduced.

A network of artificial neurons connected with each other allowed parallel signal processing distributed along various branches of the network. The connections b/w the “neuron” units contained weights to control strength of the effect a neuron has on another.



ANN from wikipedia

This approach was perceived to be quite similar to what happens inside our nervous system and this caused some hipe among the researchers about the effectiveness of these models.

During this wave of Connectionism, various models like LSTM, distributed representation and processing, back-propagation to train deep neural nets were developed and continue to remain key components of various advanced applications of Deep learning to this date.

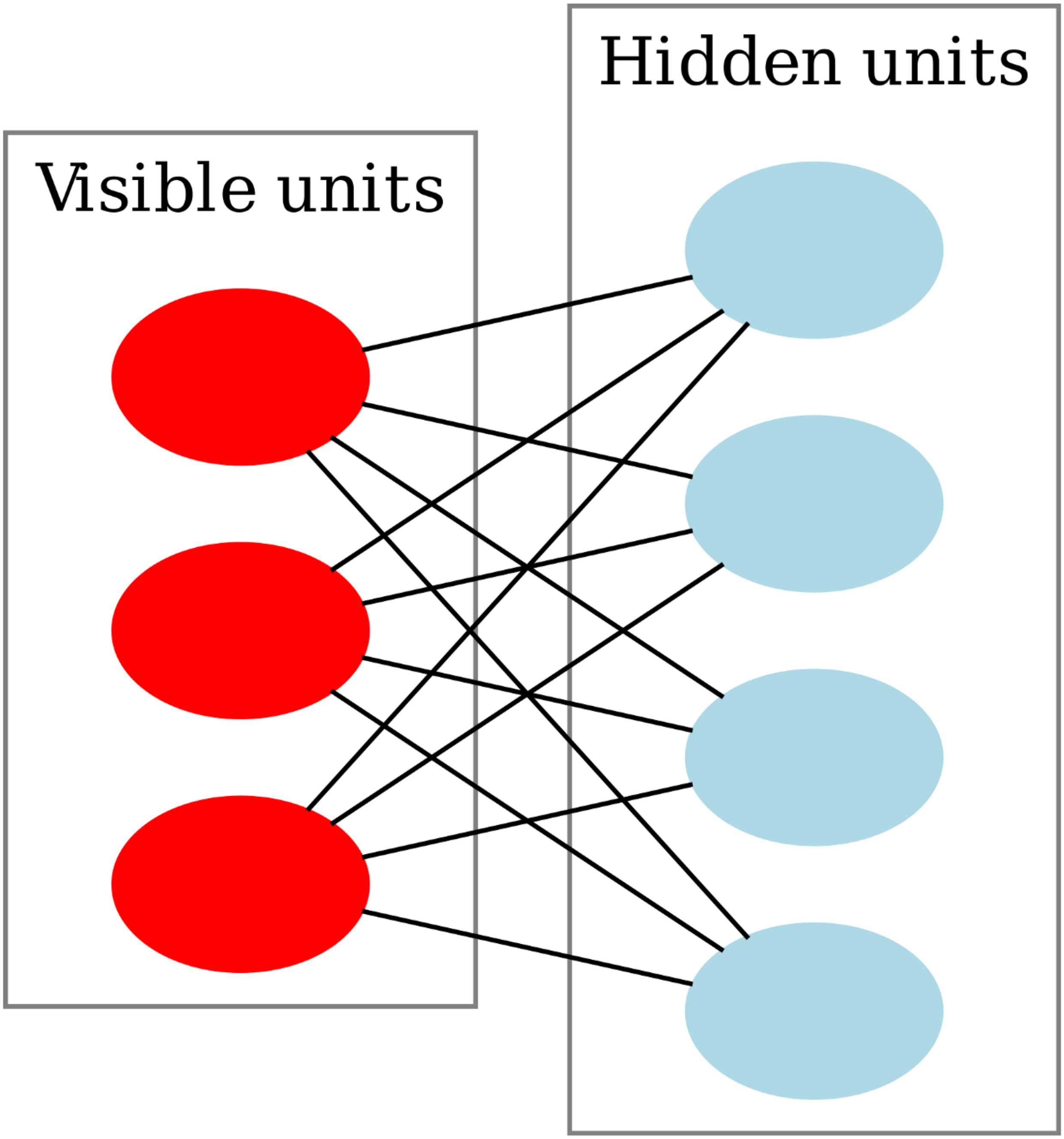
But in the mid-1990s the startups based on AI started to make unrealistic claims and could never deliver that level of sophistication from these models due to the lack of computational resources. Investors pulled back and this led to dip in this second wave of Deep Learning.

The second wave never died but was diminished. Research went on in various labs but applications were very few until early 2000s.

Deep Learning

After two dips, the third wave emerged in 2006 with a breakthrough. Geoffrey Hinton used Greedy Layer-wise Training to train Deep Belief Networks.

In the simplest of forms the DBN's are a composition of multiple hidden layers and each layer containing various latent variables. Connections exist b/w layers but not between the variables inside each layer. A very simple implementation of DBN can also be called restricted Boltzmann machines.



Restricted Boltzmann Machine from wikipedia

The advancements by Geoffrey Hinton were used by other researchers to train different types of Deep Networks. This enabled researchers around the world to trainer deeper and deeper neural networks and led to the popularisation of the term ***Deep Learning.***

While it might seem that Geoffrey Hinton lead to emergence of Deep Learning, you can't ignore the increase in computational powers and availability of large datasets.

Same algorithms developed during Connectionism started to give better results when trained on larger and larger datasets.

The difference b/w then and now, is, with more and more people using online services we have a lot more data and a lot superior computational resources to work with that data, hence increasing the accuracy for various models.

A more interesting and complex applications of Deep Learning are surfacing but are at early stages of practical use. For example, Deep Learning has been used to develop 3D map of the brain (Connectome) to help neuroscientists and cognitive scientists study the brain. Pharmaceutical companies are beginning to use Deep Learning to predict how different molecules will react and help accelerate drug developmen