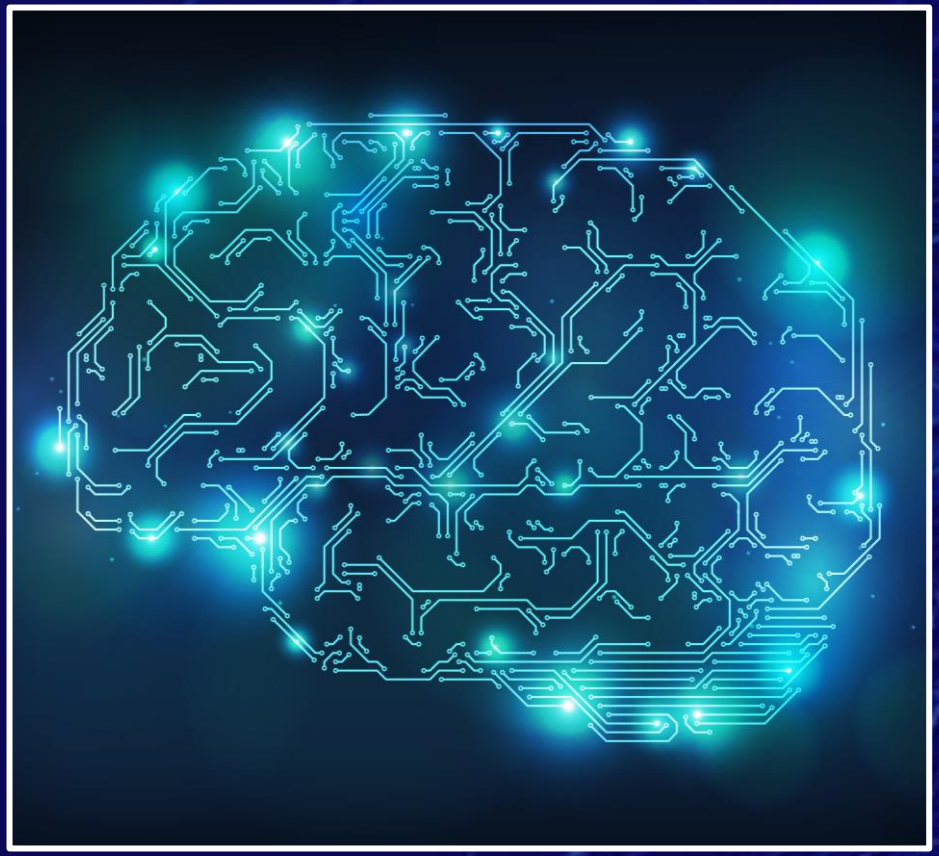
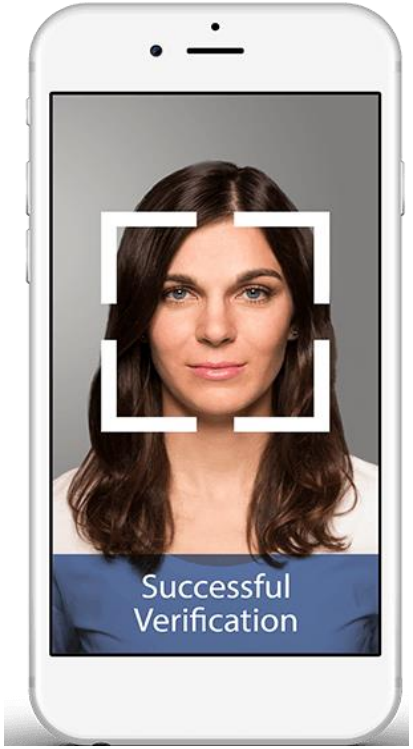
Unit 1

Fundamentals of Deep Learning



Disclaimer

The content is curated from online/offline resources and used for educational purpose only



amazon Spotify

NETFLIX Google yelp

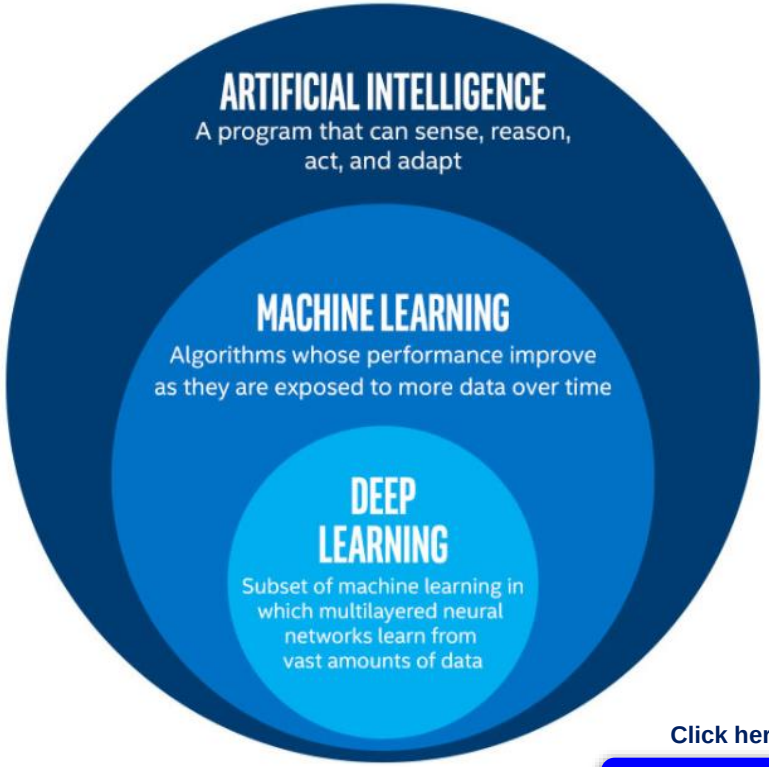
Uber Booking.com

Learning Objectives

- Deep Learning
- Deep Neural Network
- DL Vs ML
- Forward and Back Propagation
- Multilayer Perceptron

What is Deep Learning?

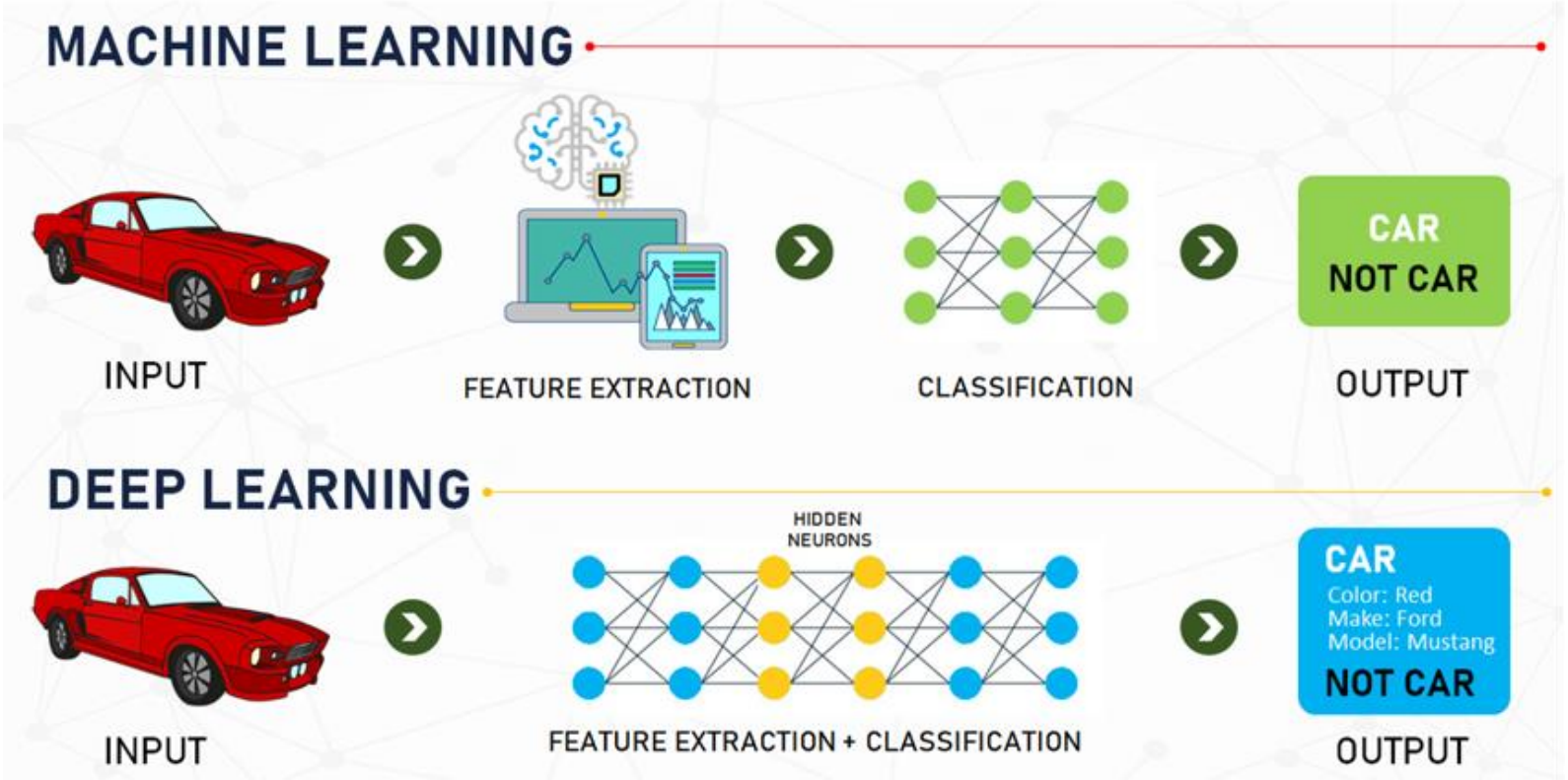
- Also known as deep structured learning
- Part of a broader family of machine learning methods
- Based on artificial neural networks
- Learning can be supervised, semi-supervised or unsupervised.



Click here

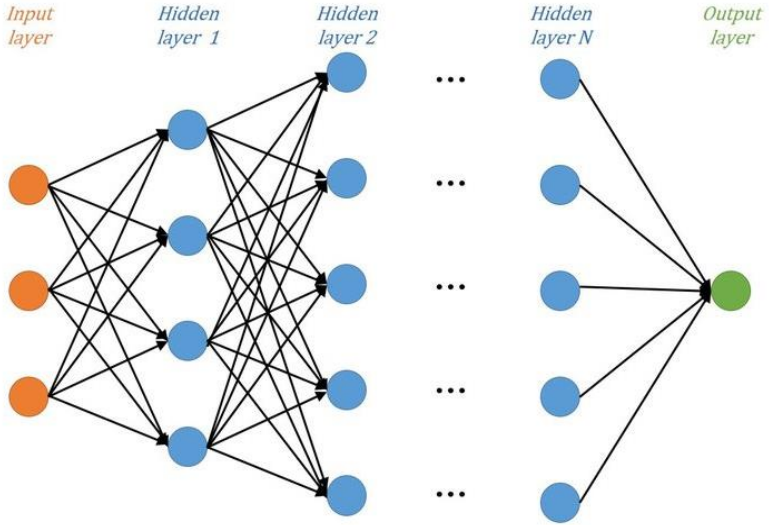
Reference link

Machine Learning Vs Deep Learning



Deep Neural Networks(DNN)

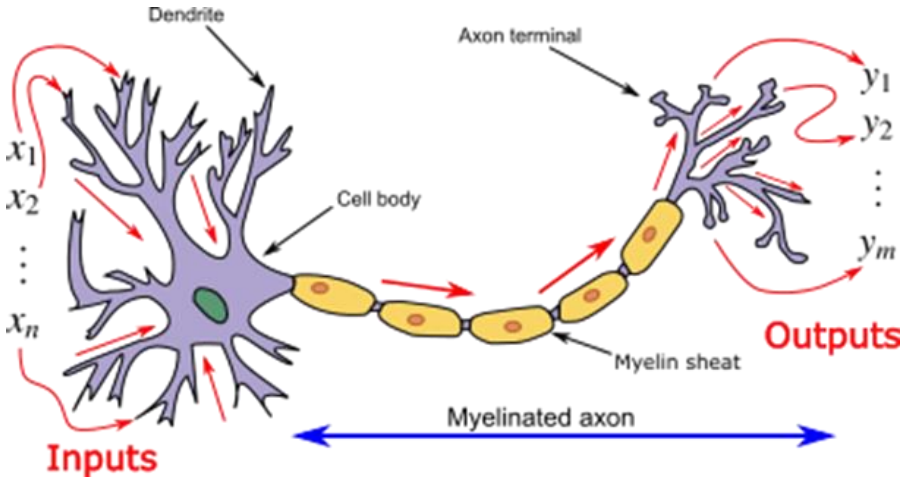
- A hierarchical organization of neurons with connections to other neurons is a simplified representation of a DNN.
- Deep Neural Networks gained their name from the fact that they utilized a lot of hidden layers to learn more intricate patterns, giving them the label "deep".
- Based on the received input, these neurons send a message or signal to other neurons, forming a complex network that learns through a feedback loop.



Neural Network Architechture

Neurons

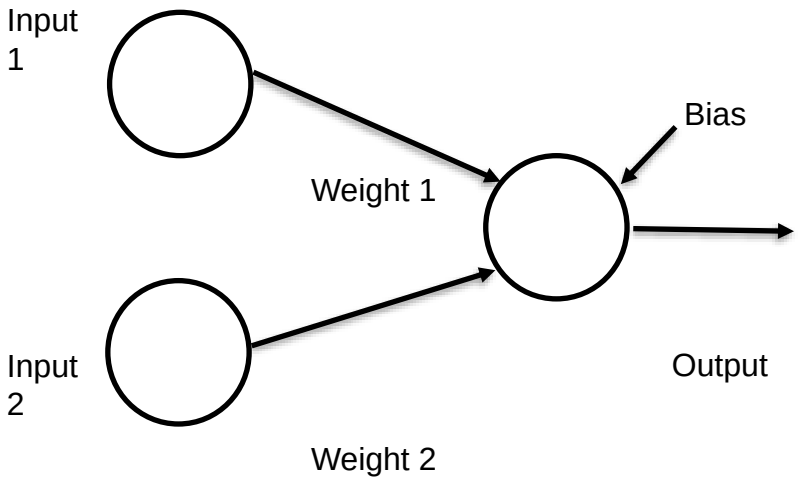
- Neurons in deep learning models are nodes through which data and computations flow.
- Neurons receive one or more input signals.
- These input signals can come from either the raw data set or from neurons positioned at a previous layer of the neural net.



Neuron

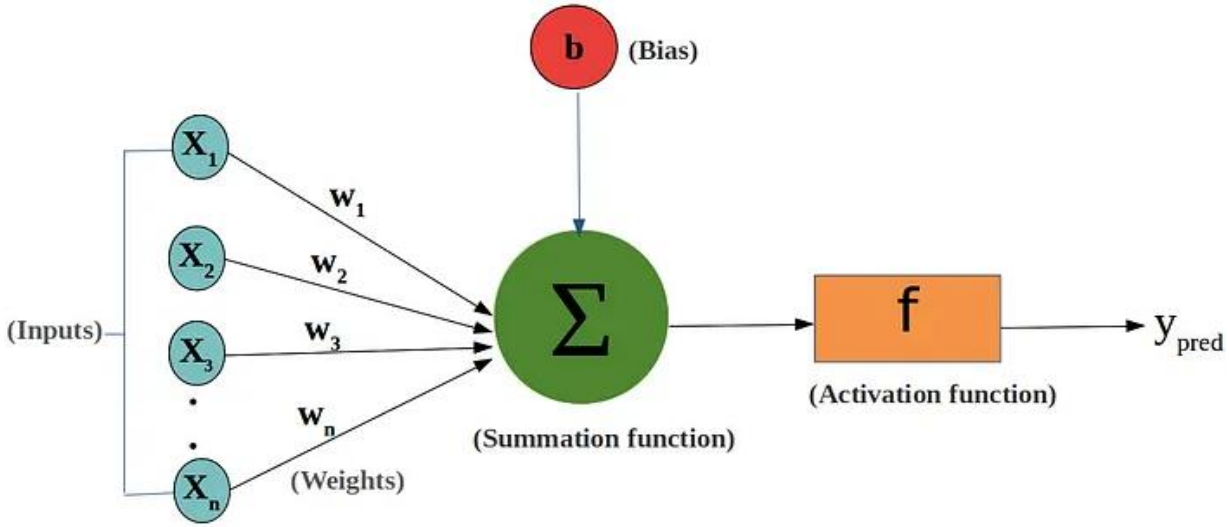
Weights and Bias

- Weights enable the artificial neural network to dial up or dial down connections between neurons.
- Bias can be used to make adjustments within neurons. Bias can be positive or negative, increasing or decreasing a neuron's output



Neuron

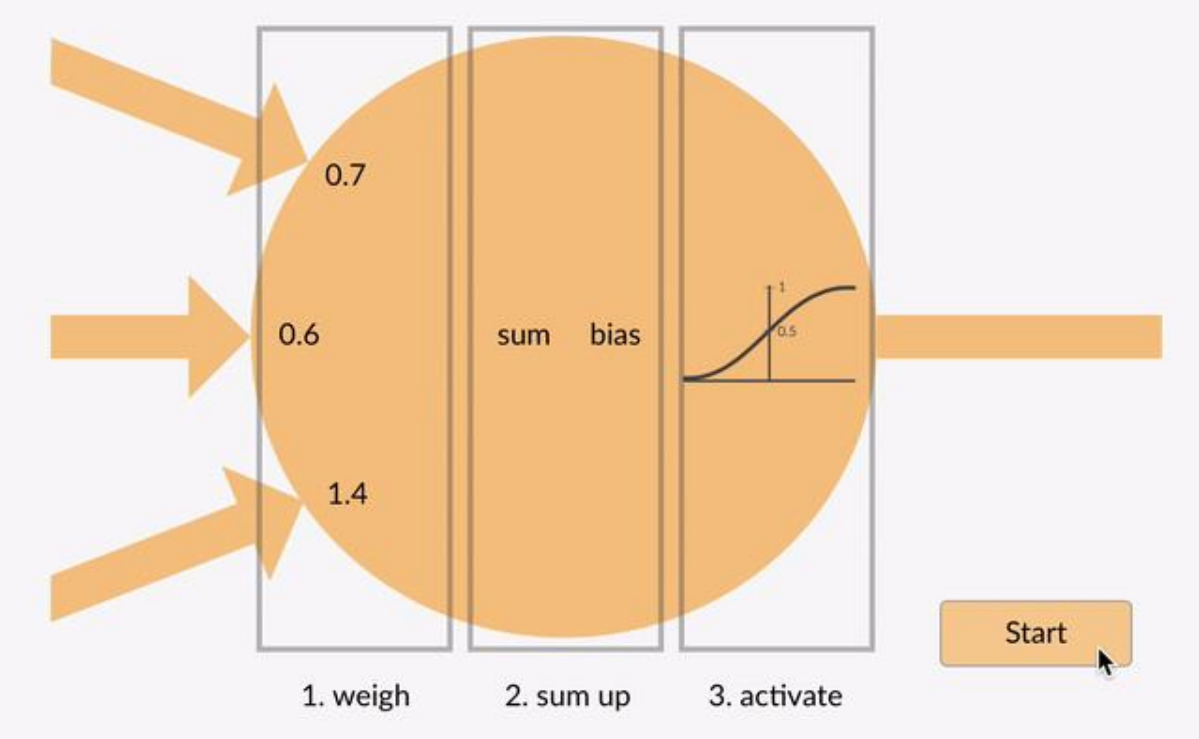
Components of Artificial Neuron



Click here

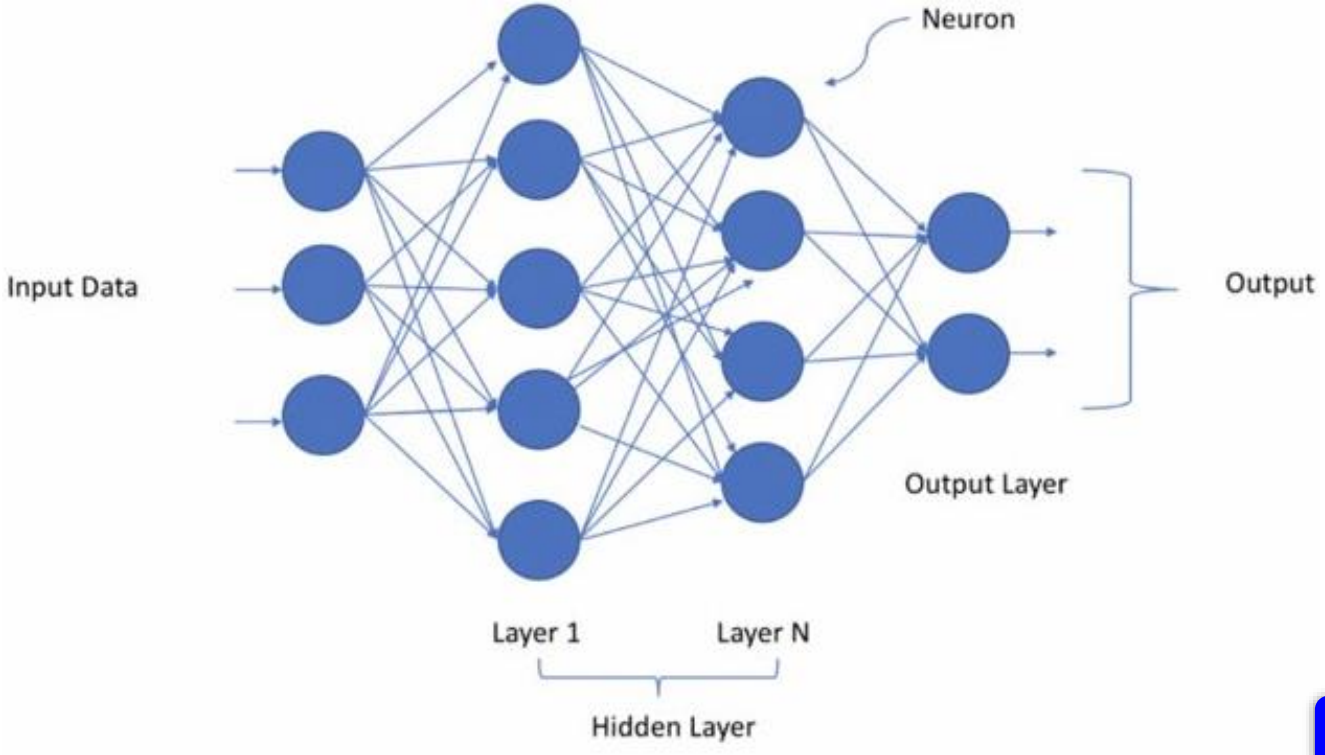
Reference link

Single Neuron



Working of single neuron

Representation of DNN



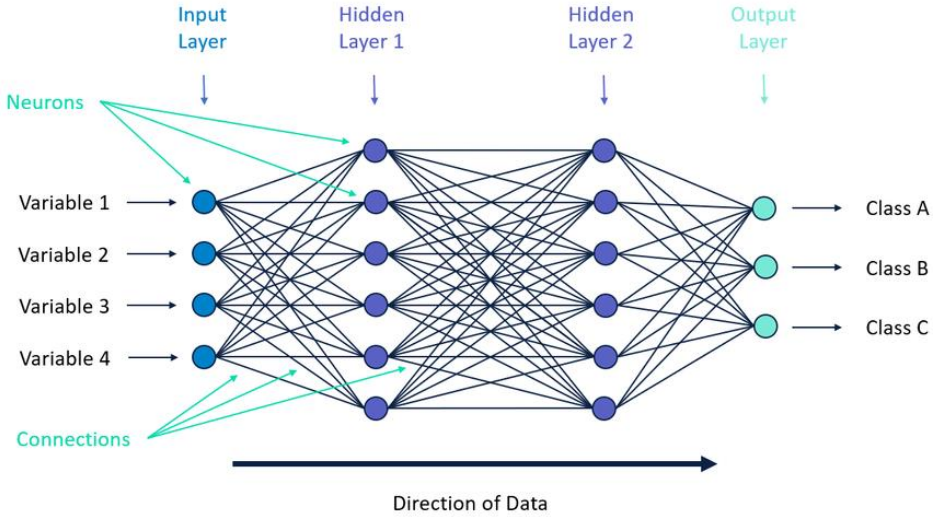
Click here

Reference link

Forward Propagation

- Forward propagation is how neural networks make predictions. Input data is “forward propagated” through the network layer by layer to the final layer which outputs a prediction.

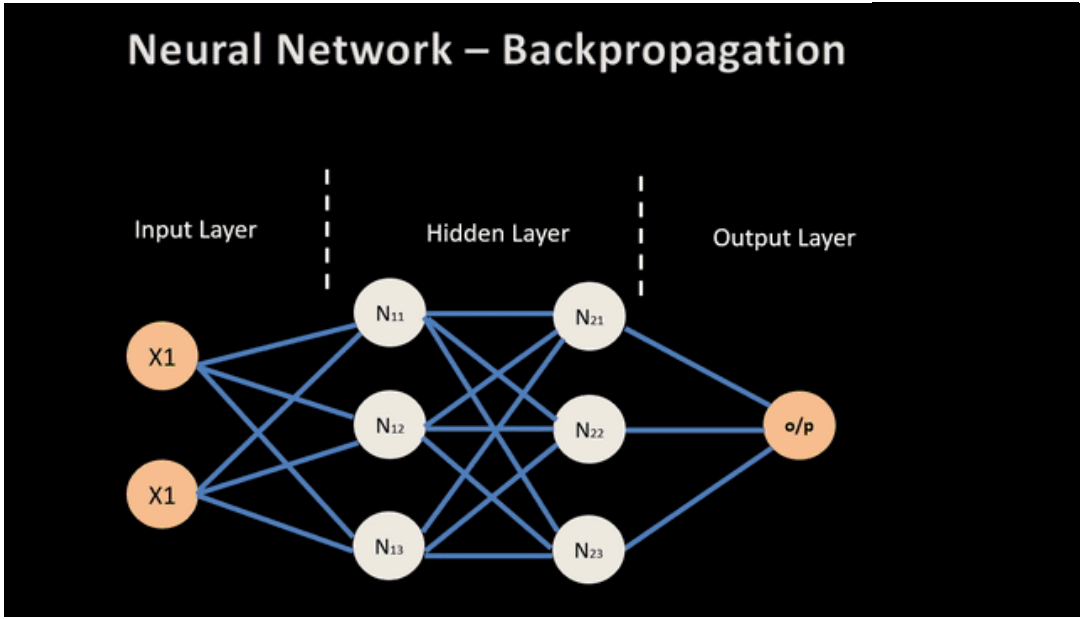
- Forward propagation (or forward pass) refers to the calculation and storage of intermediate variables (including outputs) for a neural network in order from the input layer to the output layer.



Backward Propagation

- In machine learning, backpropagation is a widely used algorithm for training feedforward neural networks. Generalizations of backpropagation exist for other artificial neural networks (ANNs), and for functions generally. These classes of algorithms are all referred to generically as "backpropagation".
- In fitting a neural network, backpropagation computes the gradient of the loss function with respect to the weights of the network for a single input–output example, and does so efficiently, unlike a naive direct computation of the gradient with respect to each weight individually.

Working of ANN

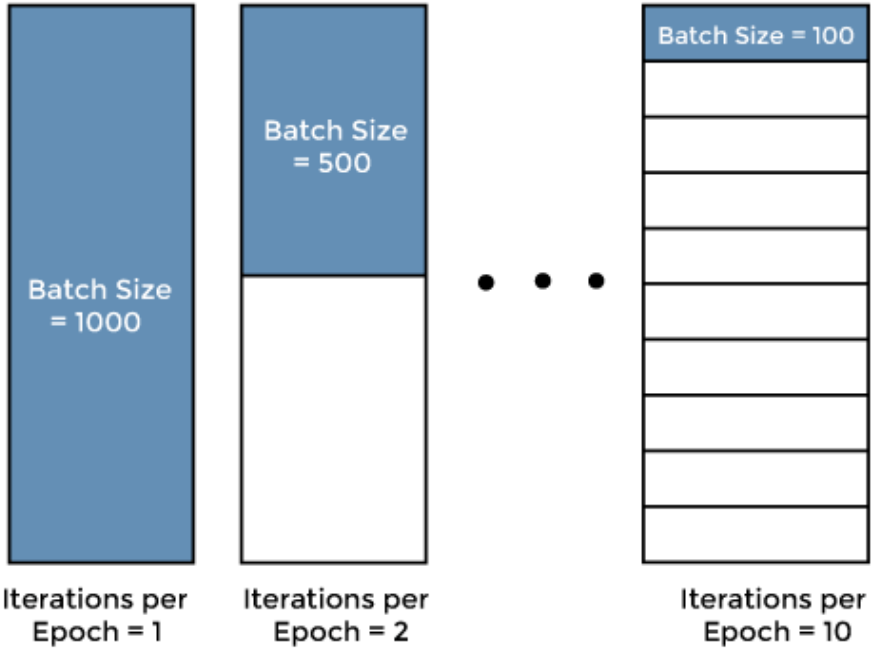


Epochs

- One Epoch is when an ENTIRE dataset is passed forward and backward through the neural network only ONCE.
- Since one epoch is too big to feed to the computer at once we divide it in several smaller batches.

Batch Size

- Total number of training examples present in a single batch.
- Note: Batch size and number of batches are two different things.



Click here

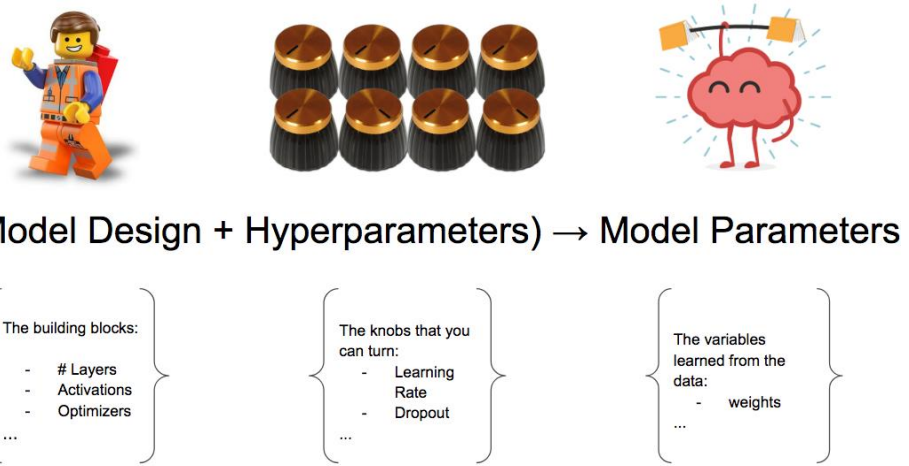
Reference link

Hyperparameters in Neural Network

- Hyperparameters are the variables.
- It determines the network structure

Example: Number of Hidden Units and Learning Rate.

- Hyperparameters are set before training(before optimizing the weights and bias).



Click here

Reference link

Hyperparameters related to Network Structure

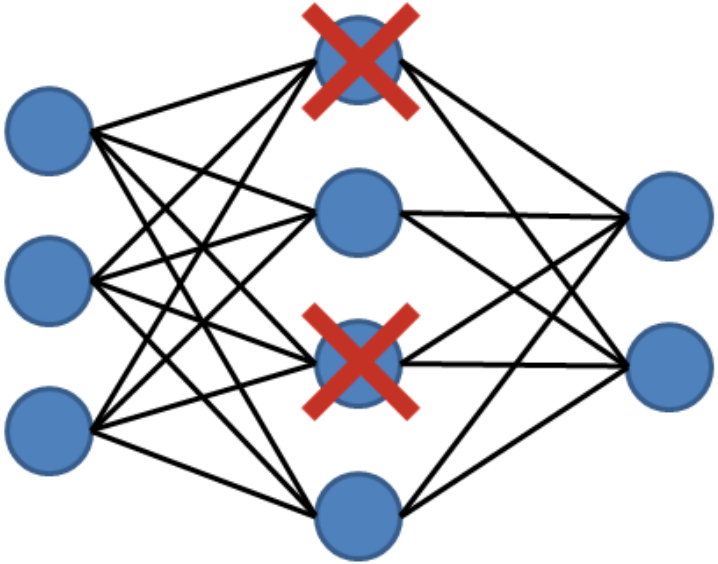
Number of Hidden Layers and Units

- Hidden layers are the layers between input layer and output layer.
- “Very simple. Just keep adding layers until the test error does not improve anymore.”
- Many hidden units within a layer with regularization techniques can increase accuracy. Smaller number of units may cause underfitting.

Hyperparameters related to Network Structure

Dropout

- Random neurons are cancelled
- Dropout is regularization technique to avoid overfitting (increase the validation accuracy) thus increasing the generalizing power.
- Generally, use a small dropout value of 20%-50% of neurons with 20% providing a good starting point. A probability too low has minimal effect and a value too high results in under-learning by the network.

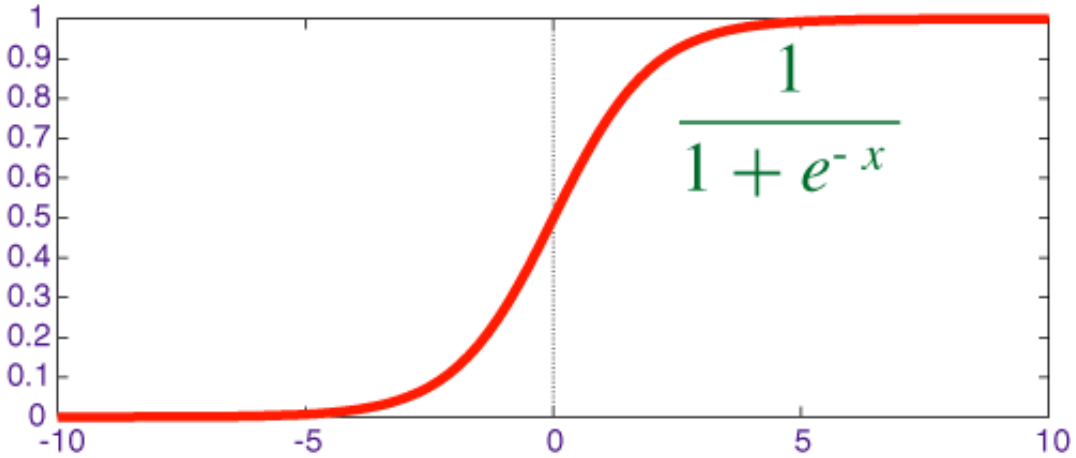


Hyperparameters related to Network Structure

Activation Function

- Activation functions are used to introduce nonlinearity to models, which allows deep learning models to learn nonlinear prediction boundaries.

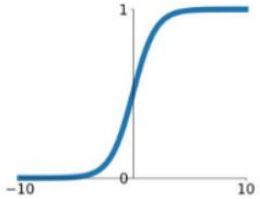
Generally, the sigmoid activation function is the most popular.



Hyperparameters related to Network Structure

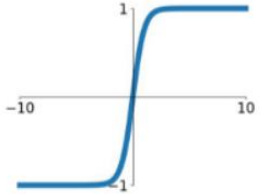
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



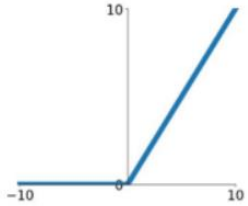
tanh

$$\tanh(x)$$



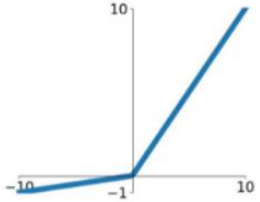
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

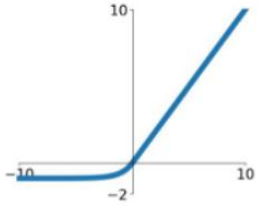


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



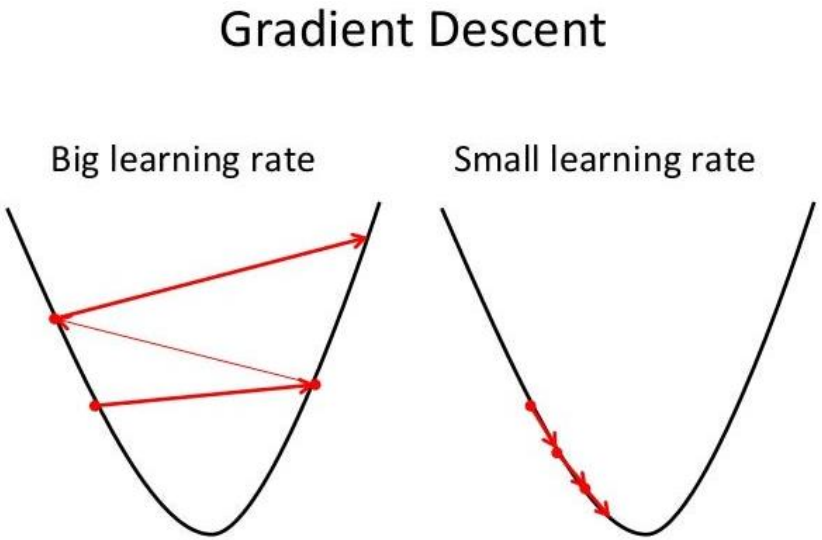
Click here

Reference link

Hyperparameters related to Training Algorithm

Learning Rate

- The learning rate defines how quickly a network updates its parameters.
- **Low learning rate** slows down the learning process but converges smoothly. **Larger learning rate** speeds up the learning but may not converge.
- Usually, a **decaying Learning rate** is preferred.



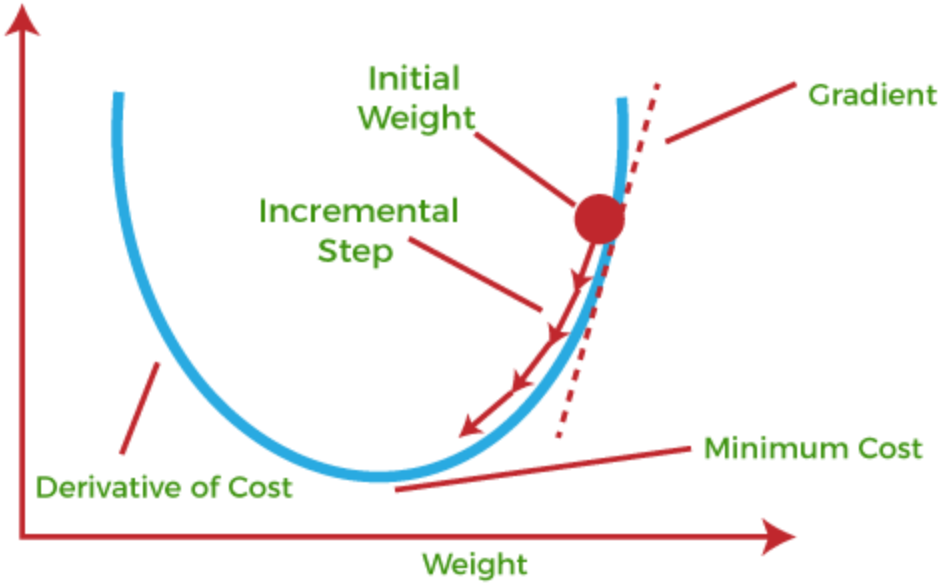
Gradient Descent

- Gradient Descent is known as one of the most used optimization algorithms to train machine learning models by means of minimizing errors between actual and expected results.
- Gradient descent was initially discovered by "Augustin-Louis Cauchy" in the mid 18th century.
- Gradient Descent is defined as one of the most used iterative optimization algorithms of machine learning to train the machine learning and deep learning models. It helps in finding the local minimum of a function.

The best way to define the local minimum or local maximum of a function using gradient descent is as follows:

- If we move towards a negative gradient or away from the gradient of the function at the current point, it will give the local minimum of that function.
- Whenever we move towards a positive gradient or towards the gradient of the function at the current point, we will get the local maximum of that function.

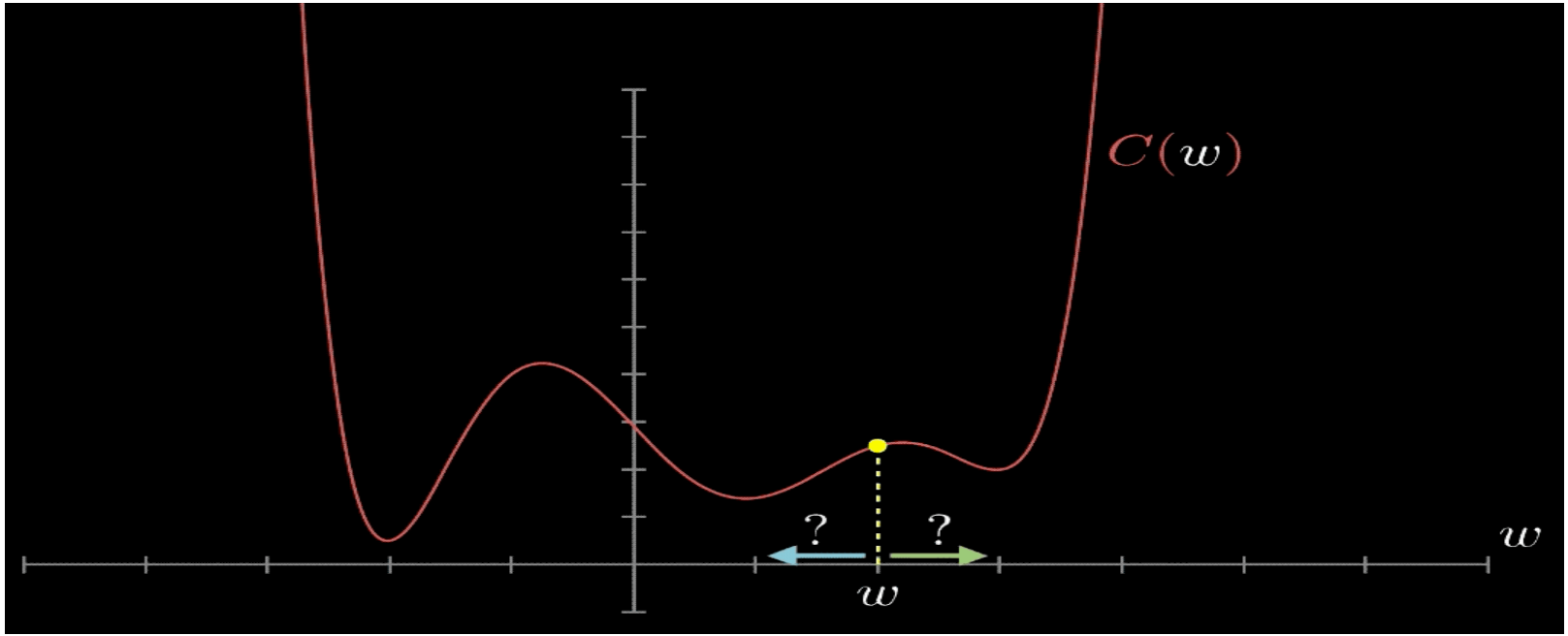
Gradient Descent



Click here

Reference link

Gradient Descent



Click here

Reference link

Hyperparameters related to Training Algorithm

Number of epochs

Number of epochs is the number of times the whole training data is shown to the network while training.

Increase the number of epochs until the validation accuracy starts decreasing even when training accuracy is increasing(overfitting).

Batch size

Mini batch size is the number of sub samples given to the network after which parameter update happens.

A good default for batch size might be 32. Also try 32, 64, 128, 256, and so on.

Loss Functions

A loss function, that can be used to estimate the loss of the model so that the weights can be updated to reduce the loss on the next evaluation.

- Regression Loss Functions
  - Mean Squared Error Loss
  - Mean Squared Logarithmic Error Loss
  - Mean Absolute Error Loss

1. Binary Classification Loss Functions

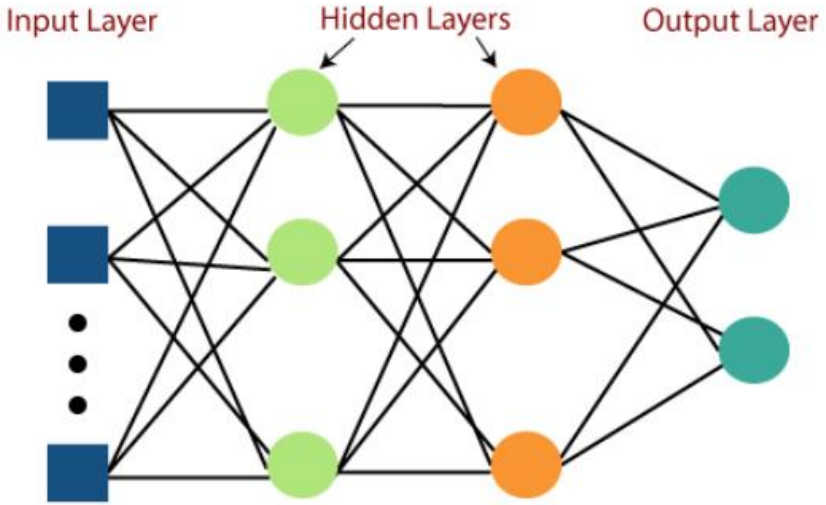
- Binary Cross-Entropy
- Hinge Loss
- Squared Hinge Loss

2. Multi-Class Classification Loss Functions

- Multi-Class Cross-Entropy Loss
- Sparse Multiclass Cross-Entropy Loss
- Kullback Leibler Divergence Loss

Multilayer Perceptron

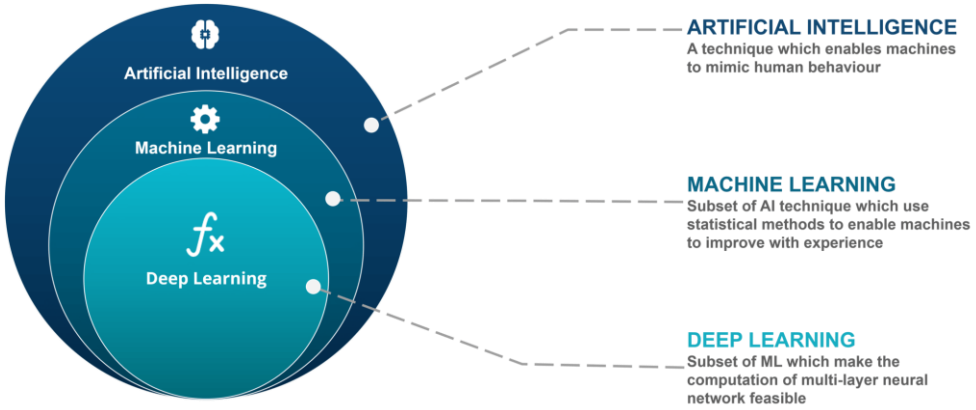
- Multilayer Perceptron is commonly used in simple regression problems. However, MLPs are not ideal for processing patterns with sequential and multidimensional data.
- MLP networks are used for supervised learning format. A typical learning algorithm for MLP networks is also called back propagation's algorithm.
- Multi-Layer perceptron defines the most complex architecture of artificial neural networks. It is substantially formed from multiple layers of the perceptron.



Multilayer Perceptron Definition | DeepAI

Advantages of DL over ML

- DL models do not require any pre-processing for feature extraction and are capable of categorizing data into multiple classes and categories on their own.
- Raw Data is given as an input to DL model.
- Pre-processed data is given to ML model.
- After a certain point, the accuracy of ML models stops rising with more data, but the accuracy of DL models continues to increase with increasing data.



Click here

Reference link

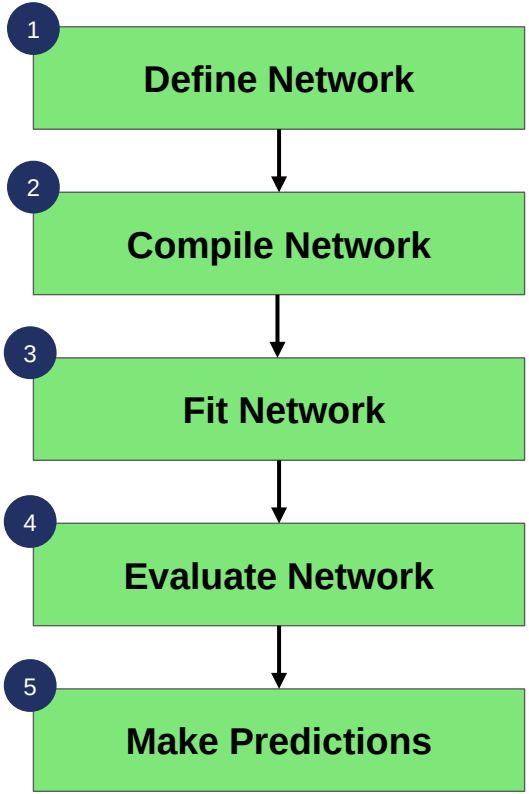
Neural Network using Keras

TensorFlow 2.0 & Keras API

Tensor

- TensorFlow 2.0 is a library that provides a comprehensive ecosystem of tools for developers, researchers, and organizations who want to build scalable Machine Learning and Deep Learning applications.
- **TensorFlow Applications**

Building a model in Keras



Build ANN using Keras

What is Keras?

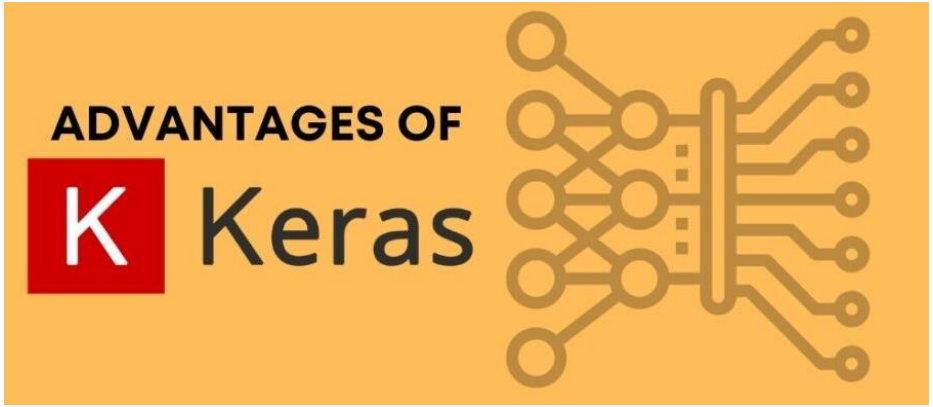
- Keras is a deep learning API written in Python, running on top of the machine learning platform TensorFlow.
- It was developed with a focus on enabling fast experimentation. Being able to go from idea to result as fast as possible is key to doing good research.

Click here

Reference link

Why Keras?

- **Simple** - but not simplistic. Keras reduces developer cognitive load to free you to focus on the parts of the problem that really matter.
- **Flexible** - Keras adopts the principle of progressive disclosure of complexity: simple workflows should be quick and easy, while arbitrarily advanced workflows should be possible via a clear path that builds upon what you've already learned.
- **Powerful -** Keras provides industry-strength performance and scalability: it is used by organizations and companies including NASA, YouTube, or Waymo.



Click here

Reference link

Keras Pipeline

Lab 1 First neural network with keras

SUMMARY

- So we have Learned what is Deep Learning , and how it is related to AI and how it is different from AI and ML
- What is deep Neural Network, neurons and components of Artificial neuron
- Forward and Backward Propagation
- Then we have learned Hyperparameter related to Network structure
- Advantage of DL over ML
- And at last Neural network using keras
-

QUIZ

1. Which of the following is not a common activation function used in deep neural networks?

A. ReLU (Rectified Linear Activation)
B. Sigmoid
C. Tanh (Hyperbolic Tangent)
D. Mean Squared Error

QUIZ

2. In deep learning, what is backpropagation used for?

A. Initializing neural network weights
B. Calculating the loss function
C. Updating neural network weights
D. Determining the number of hidden layers

QUIZ

3. What is an epoch in the context of training deep neural networks?

A. A single forward pass through the network

B. A single backward pass through the network

C. One complete iteration through the training dataset

D. The number of layers in the neural network

QUIZ

4. Which deep learning technique is used to prevent the vanishing gradient problem in very

A. Regularization
B. Batch normalization
C. Data augmentation
D. Max-pooling

QUIZ

5. What is the primary objective of deep learning?

A. Feature engineering

B. Data visualization

C. Automatic feature learning

D. Data preprocessing

What is Deep Learning?

- https://www.javatpoint.com/deep-learning-algorithms
- https://www.guru99.com/deep-learning-tutorial.html
- https://www.tutorialspoint.com/python_deep_learning/index.htm
- https://www.datacamp.com/tutorial/tutorial-deep-learning-tutorial
- https://towardsdatascience.com/what-are-hyperparameters-and-how-to-tune-the-hyperparameters-in-a-deep-neural-network-d0604917584a

Thank you...!