



DEMYSTIFYING DEEP LEARNING: TUTORIAL SERIES

CHAPTER 2: BUILDING BLOCKS OF NEURAL NETWORK-1 *Vaibhav Verdhan*

May 04 2020

AGENDA FOR THE SERIES

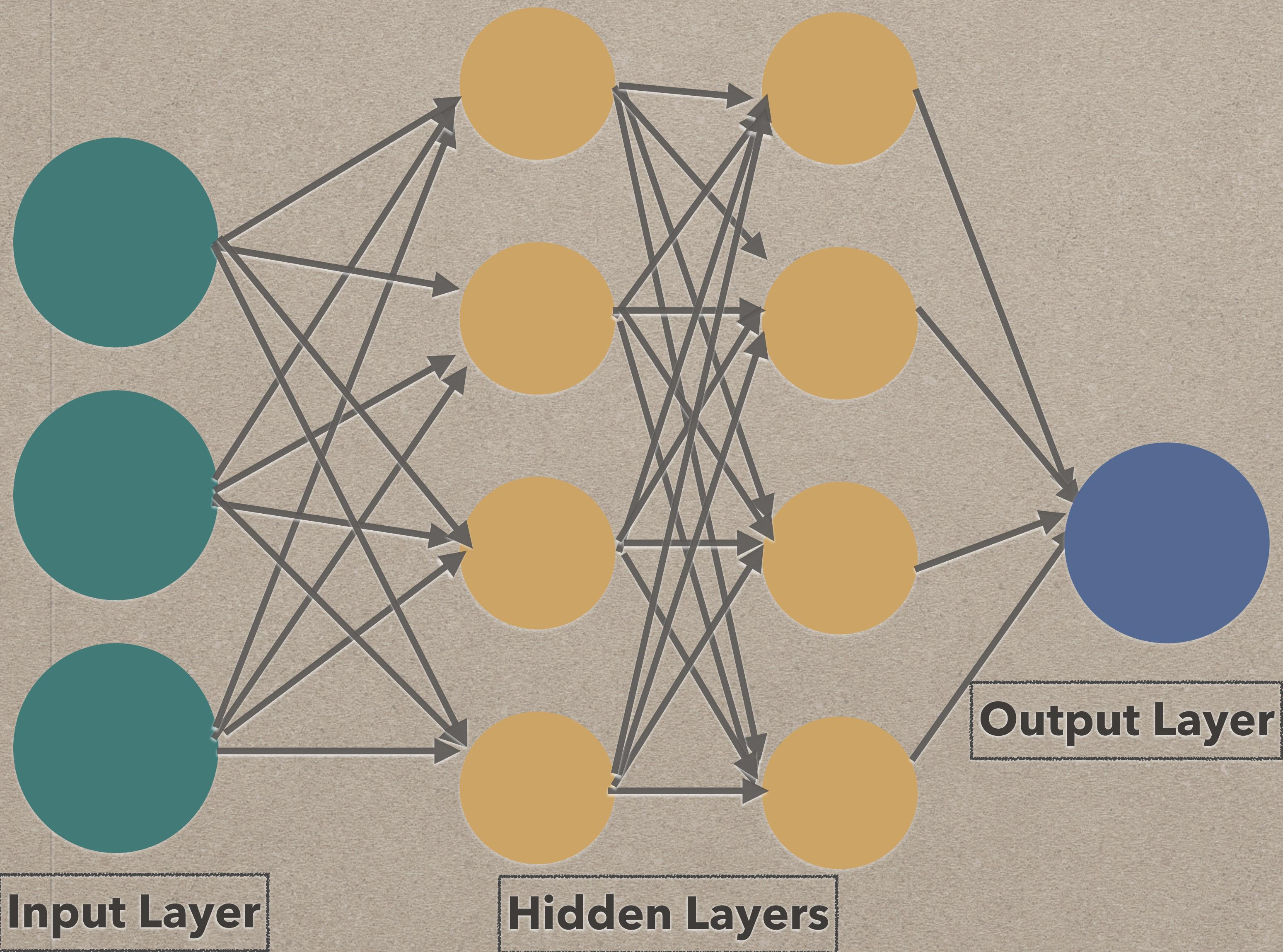
Session	Topic
Session 1	Introduction to Deep Learning
Session 2	Building blocks of Neural Network - 1
Session 3	Building blocks of Neural Network - 2
Session 4	Convolutional Neural Network
Session 5	CNN for Image Classification
Session 6	CNN for Object Detection
Session 7	Architectures like AlexNet, Inception etc.
Session 8	Recurrent Neural Network
Session 9	NLP Applications of RNN

AGENDA FOR SESSION 2

- *Components of Neural Network*
- *Process of training of a network*
- *Activation functions*
- *Feedforward Neural Network*
- *Back-propagation and loss function*

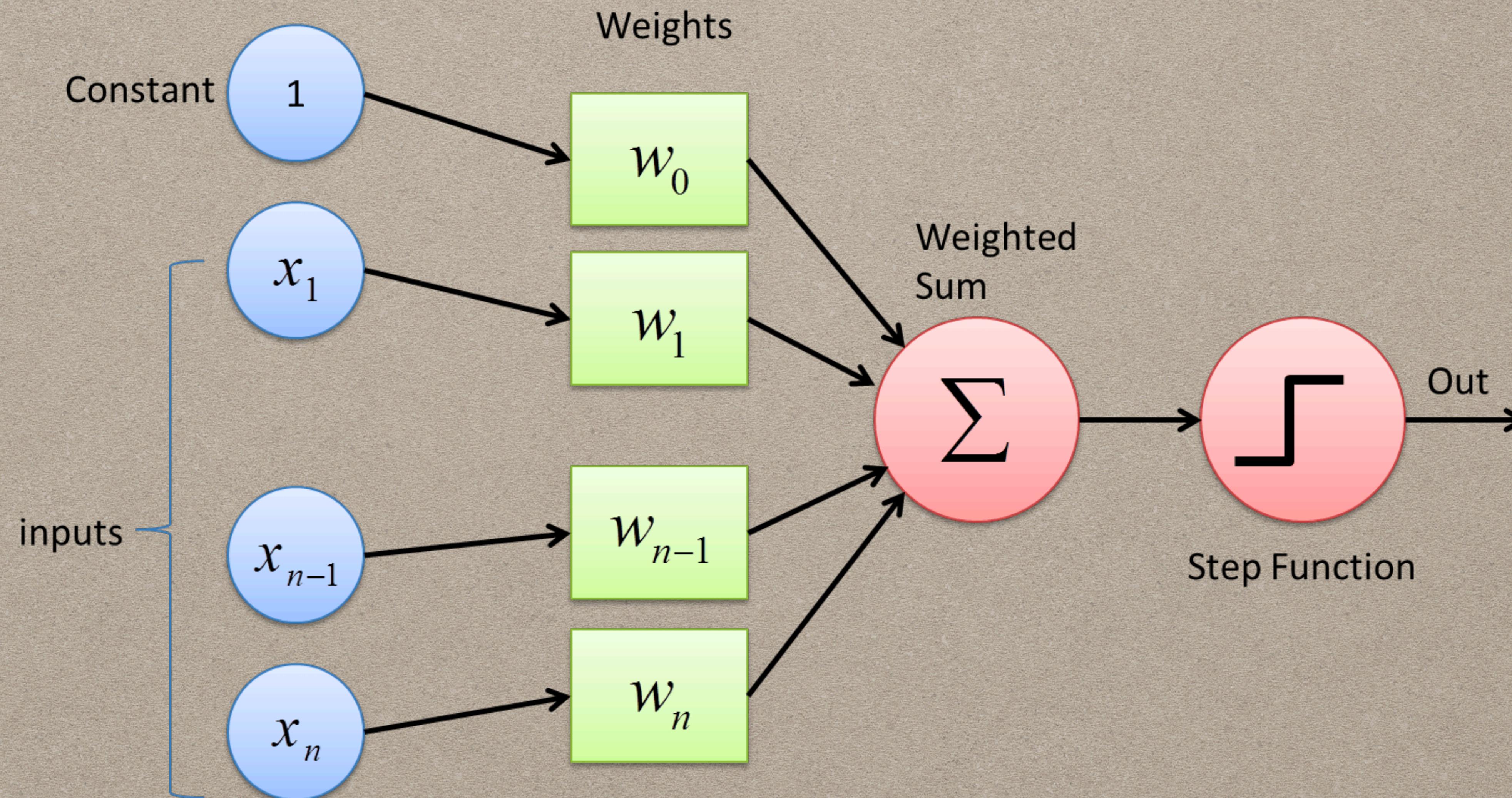


A TYPICAL NEURAL NETWORK

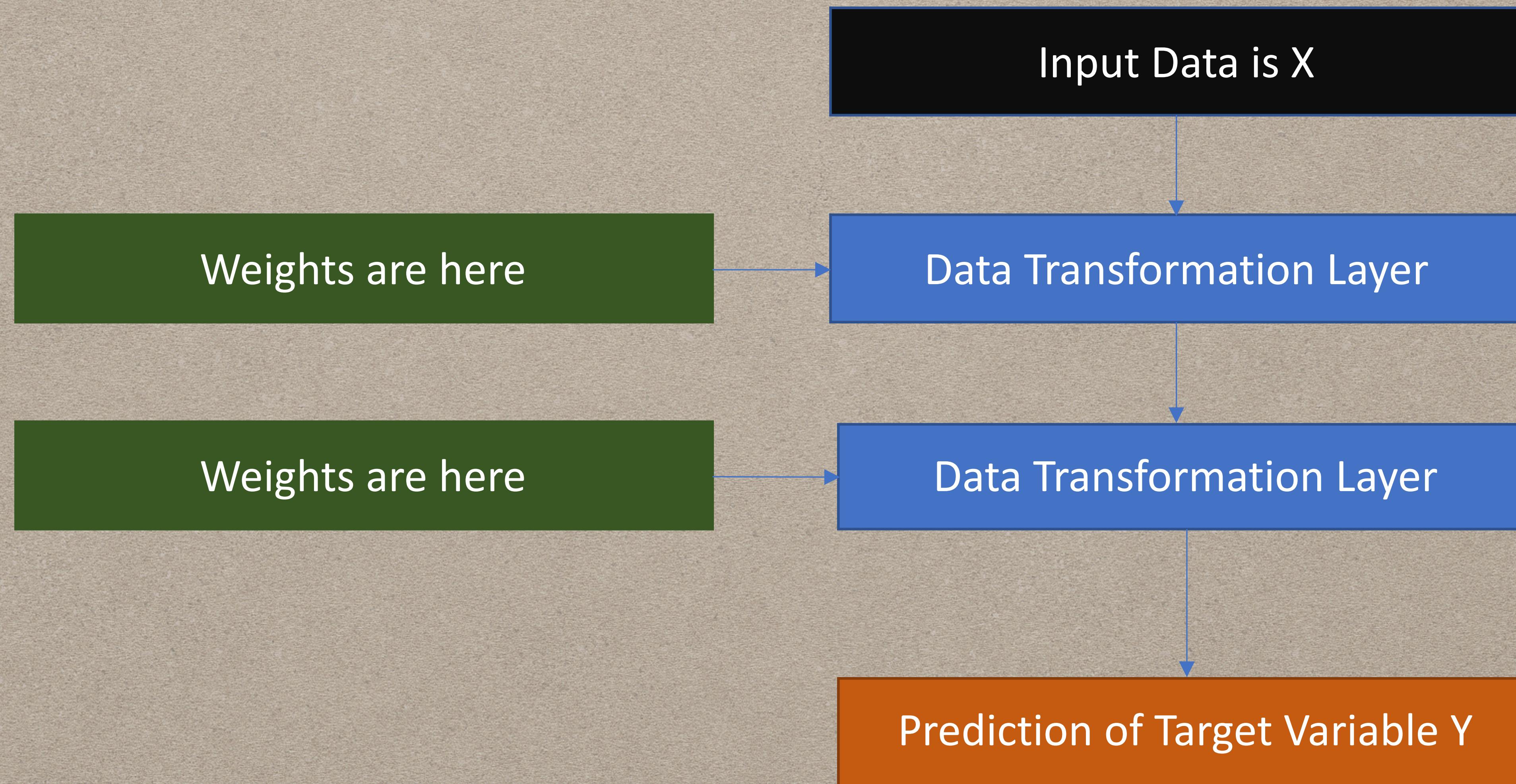


- Deep refers to the depth in the Neural Network which means the number of hidden layers
- Number of neurons in a layer are also important
- But if the network is deep, it does not necessarily mean that the accuracy will be higher

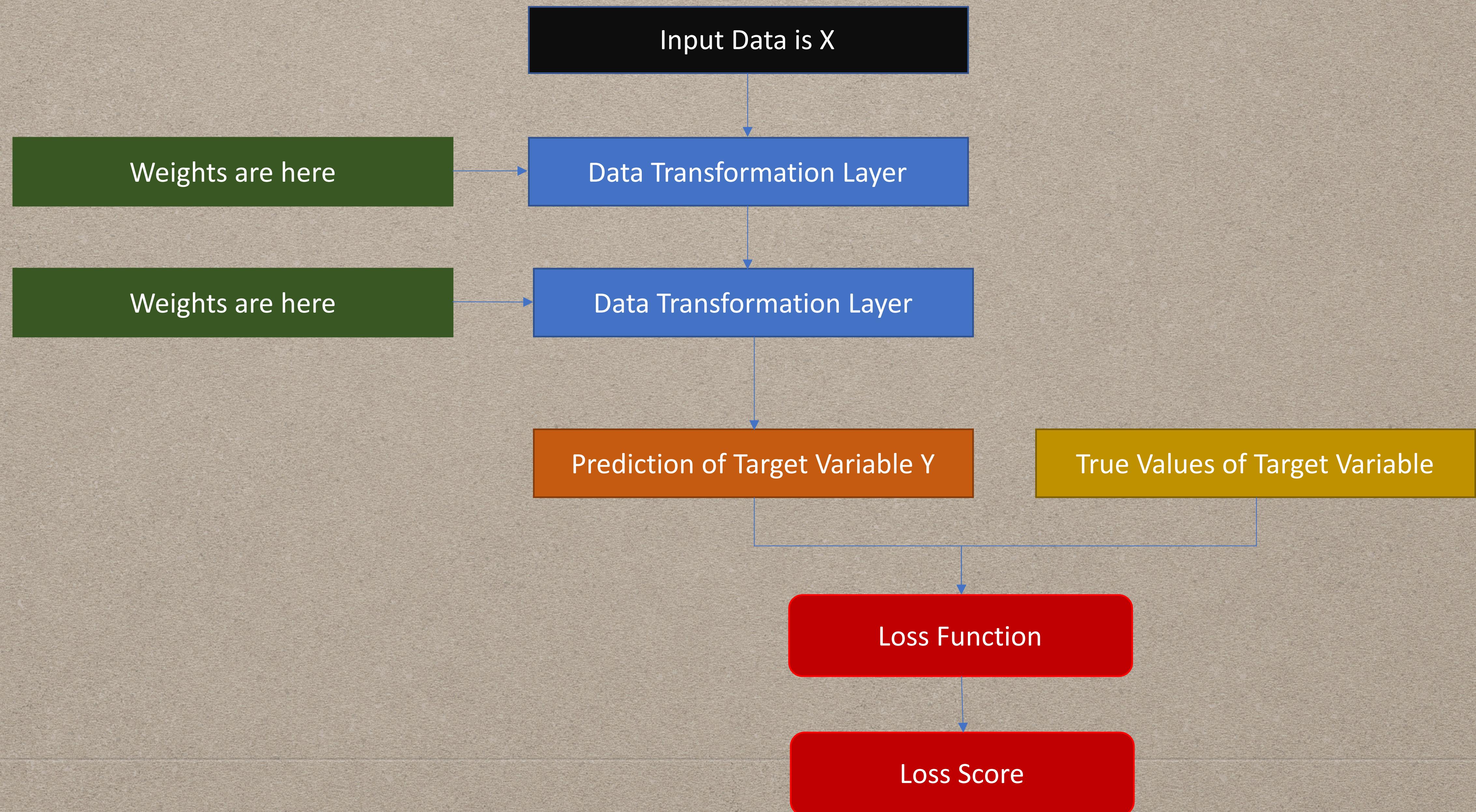
WHAT IS A PERCEPTRON



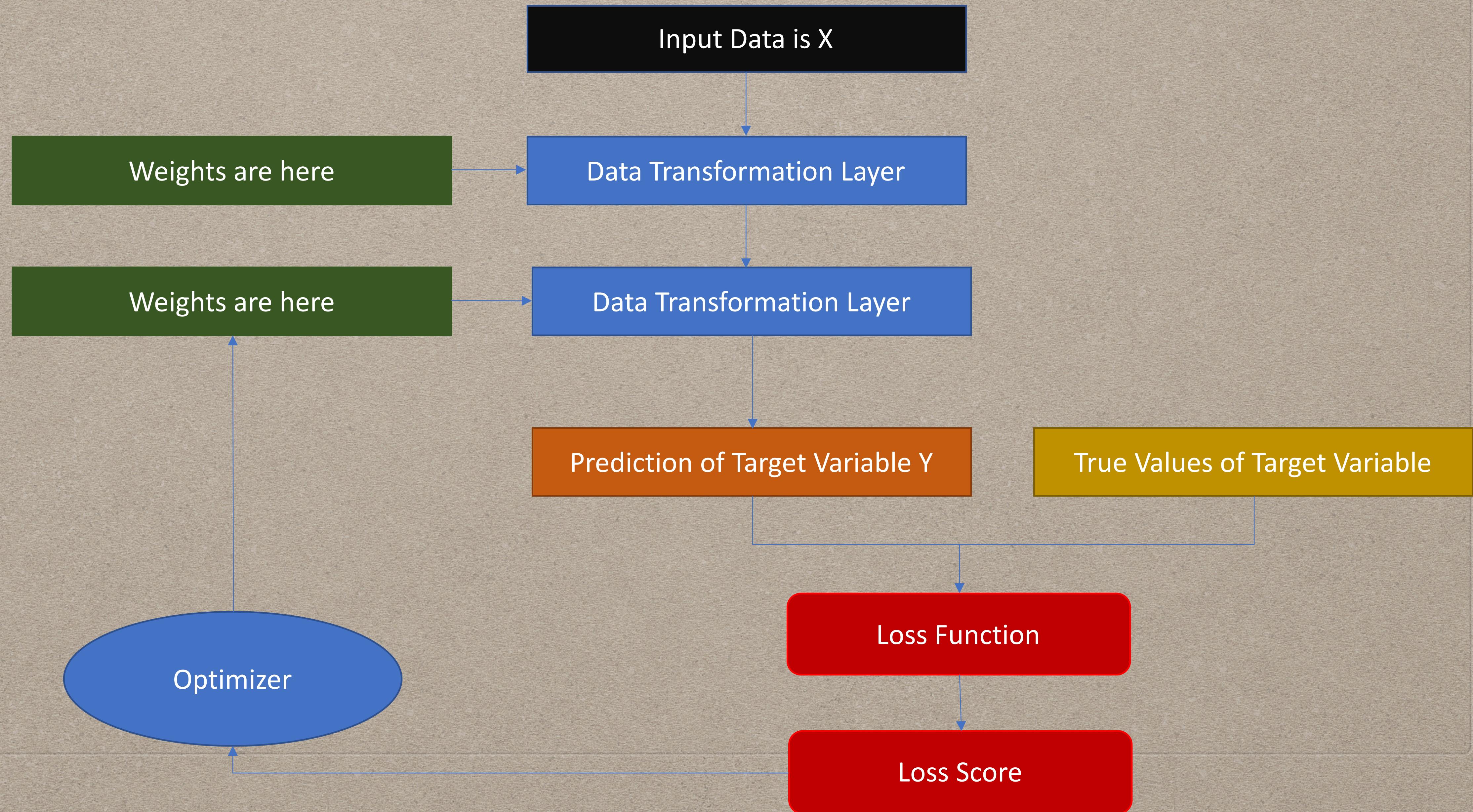
TRAINING OF A NETWORK



TRAINING OF A NETWORK...



TRAINING OF A NETWORK...

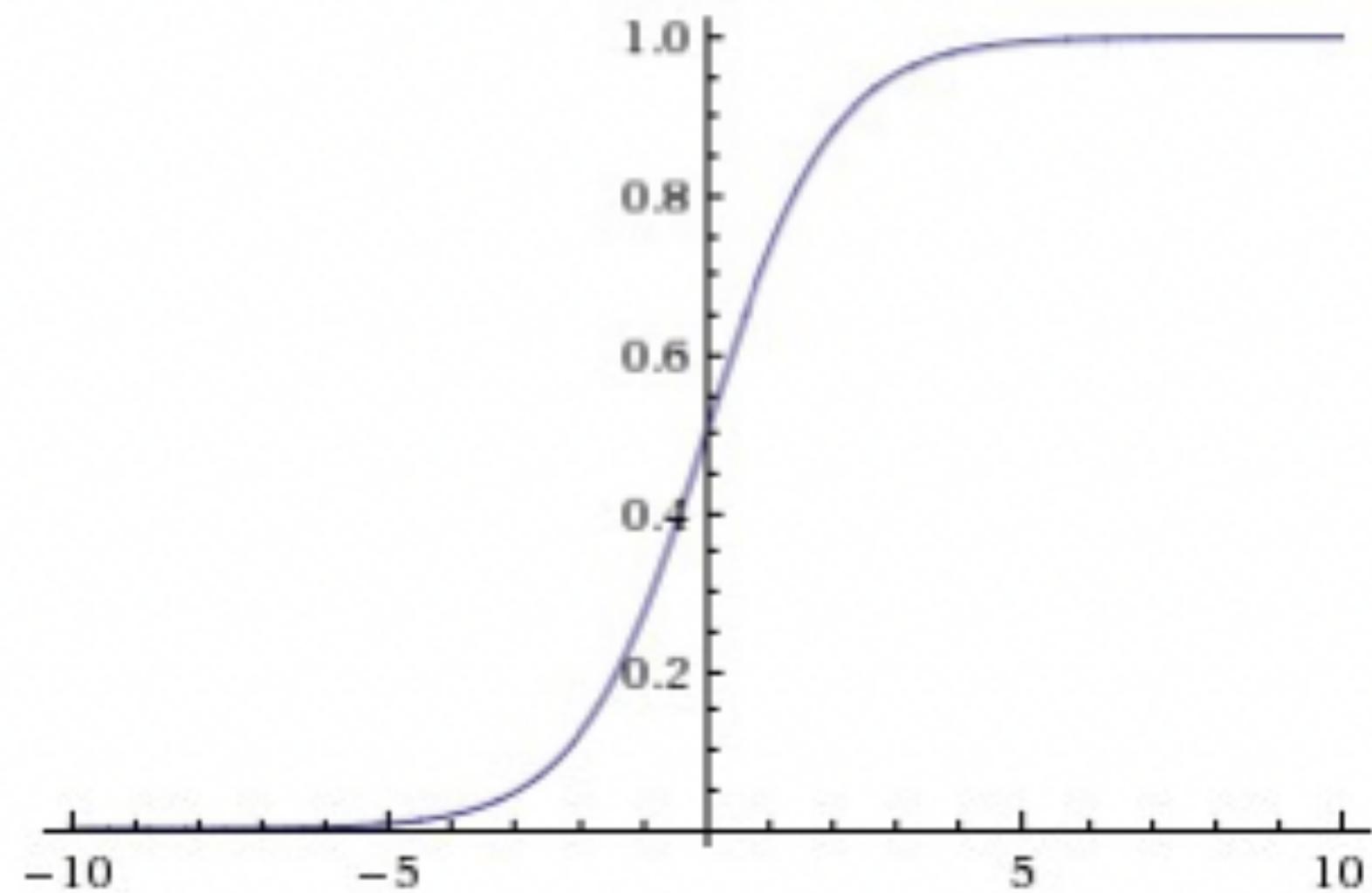


FORWARD AND BACKWARD PASS

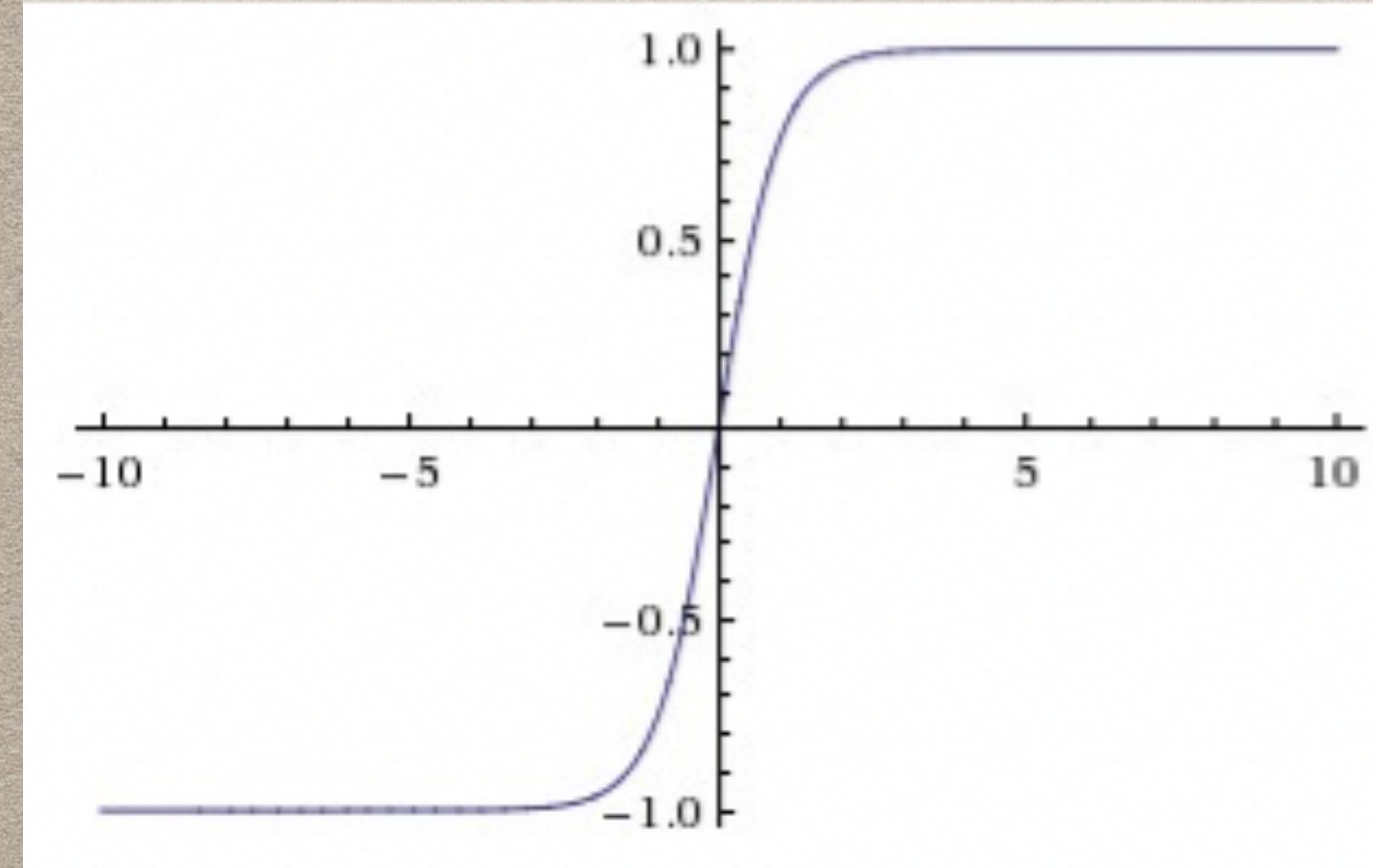
Forward pass is basically a set of operations which transform network input into the output space. During the inference stage neural network relies solely on the forward pass.

Backpropagation is an algorithm which calculates error gradients with respect to each network variable (neuron weights and biases). Those gradients are later used in optimization algorithms, such as Gradient Descent, which updates them correspondingly. The process of weights and biases update is called Backward Pass.

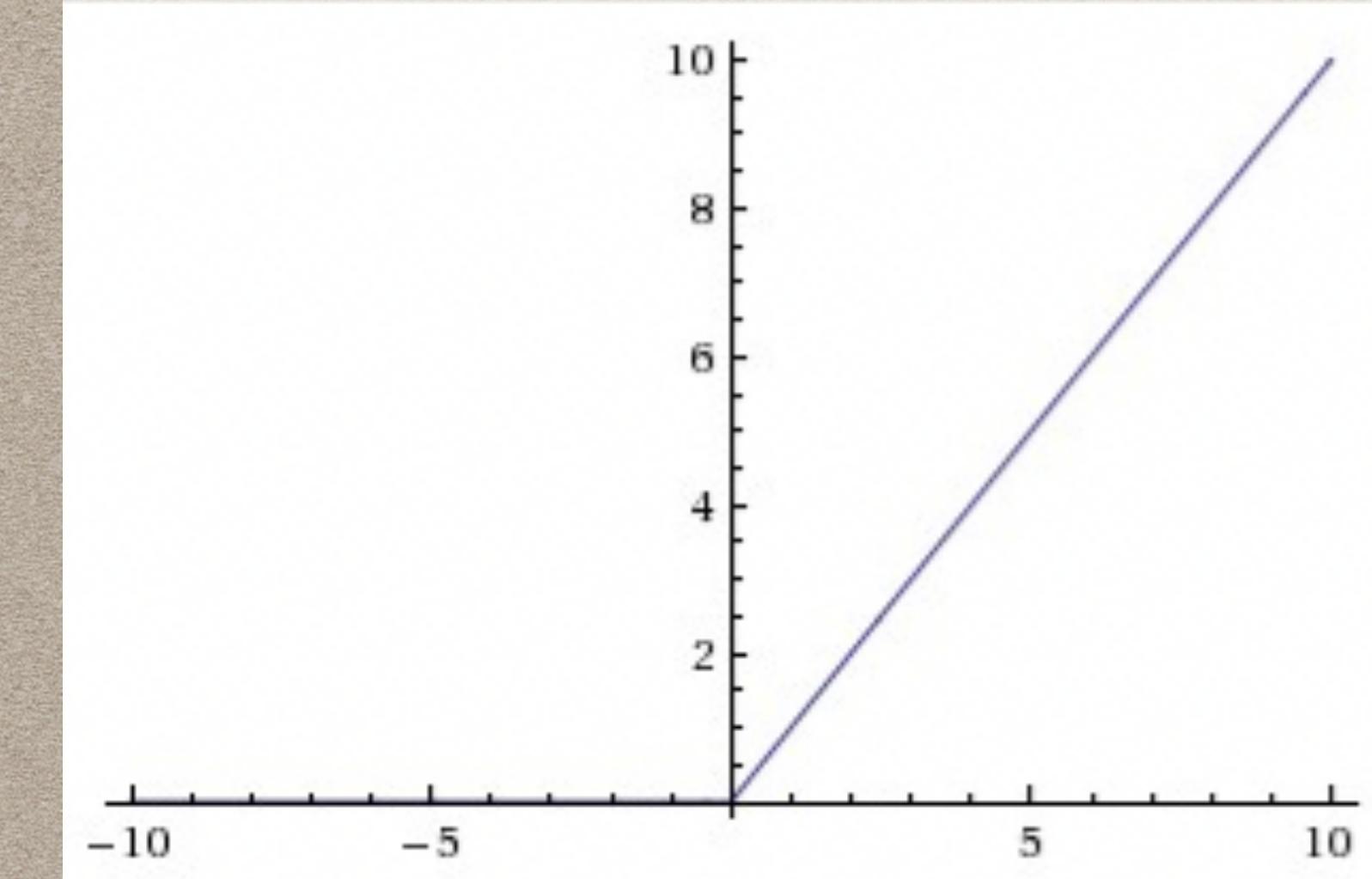
ACTIVATION FUNCTIONS



Sigmoid Function



tanh Function



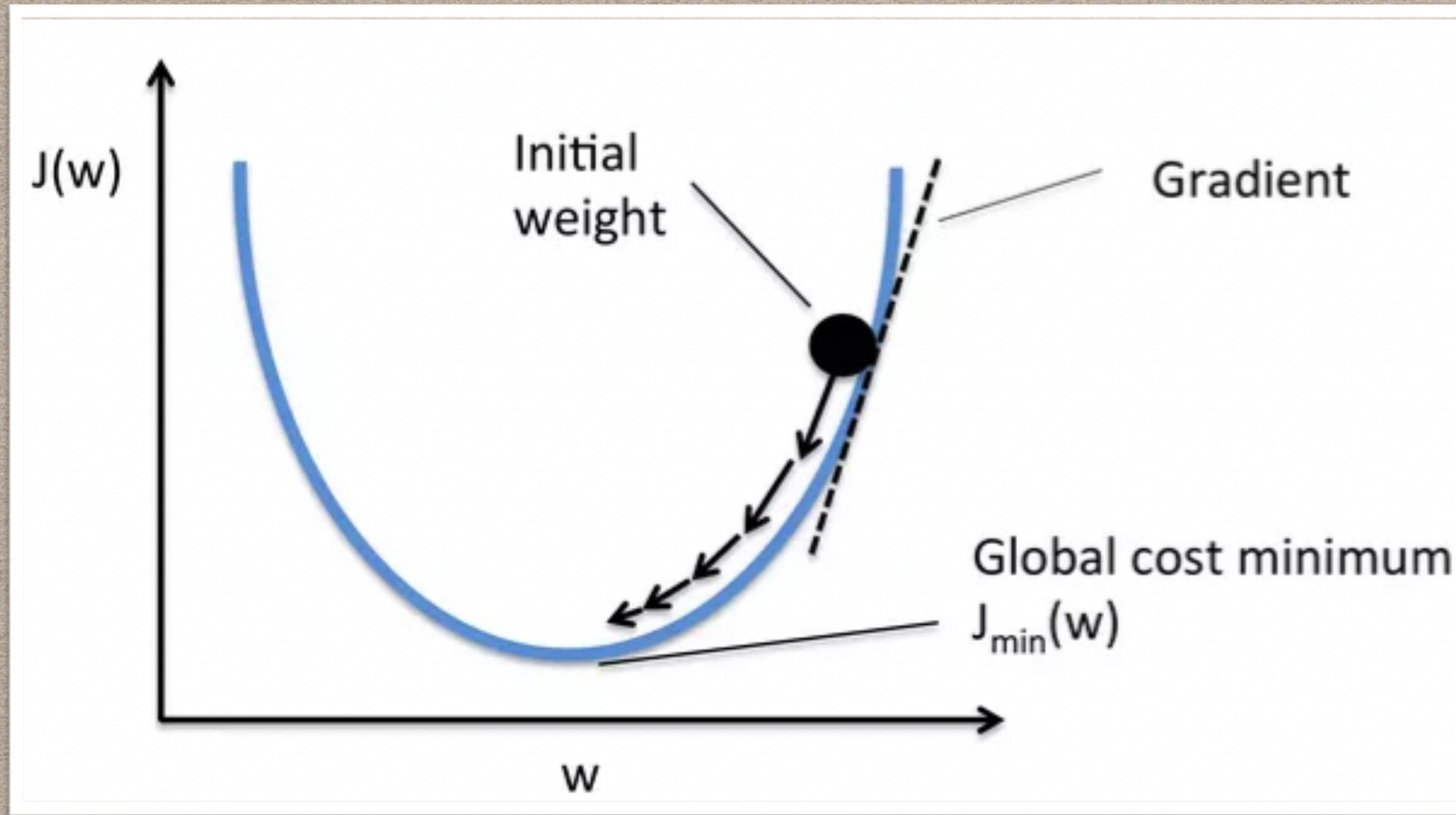
ReLU Function

- Why do we need non-linear functions?
- Every activation function (or non-linearity) takes a single number and performs certain fixed mathematical operation on it.

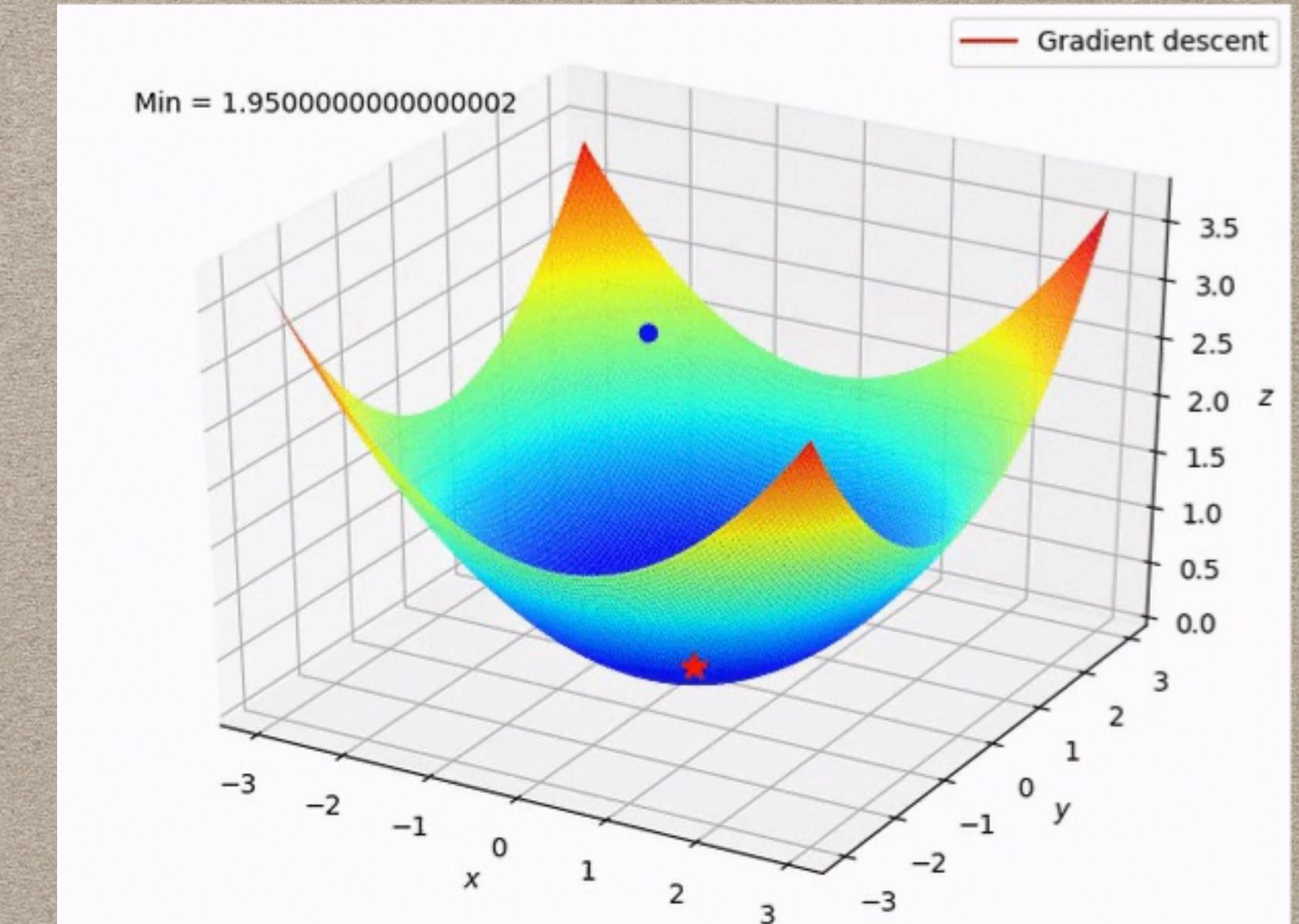
LOSS FUNCTIONS

- Error = difference between the actual value and the predicted output.
- The function that is used to compute this error is known as Loss Function.
- Different loss functions will give different errors for the same prediction, and thus have a considerable effect on the performance of the model. And hence it is advisable to have different loss functions for regression vs classification
- One of the most widely used loss function is mean square error, which calculates the square of difference between actual value and predicted value.

OPTIMISATION: FIND WEIGHTS TO MINIMISE LOSS



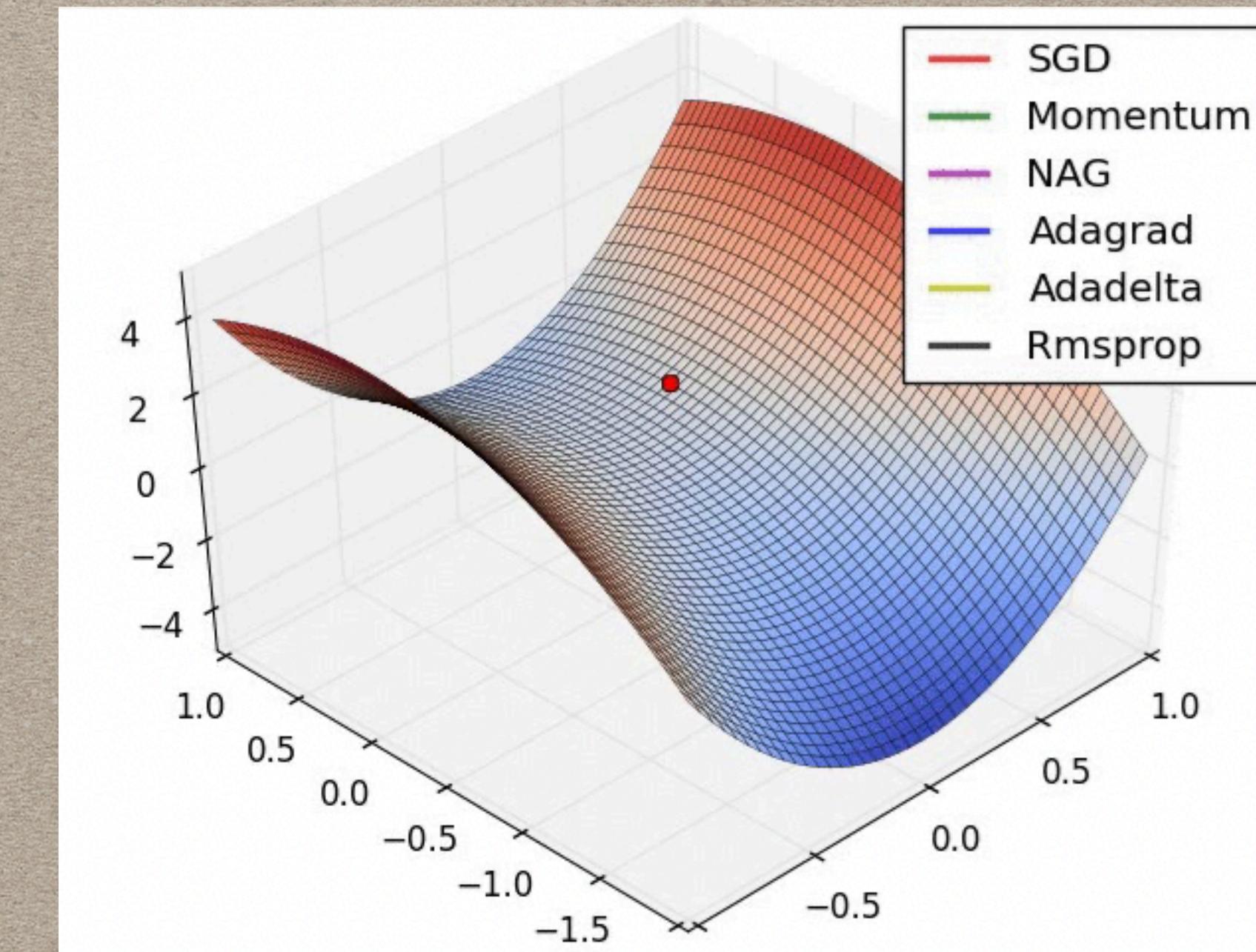
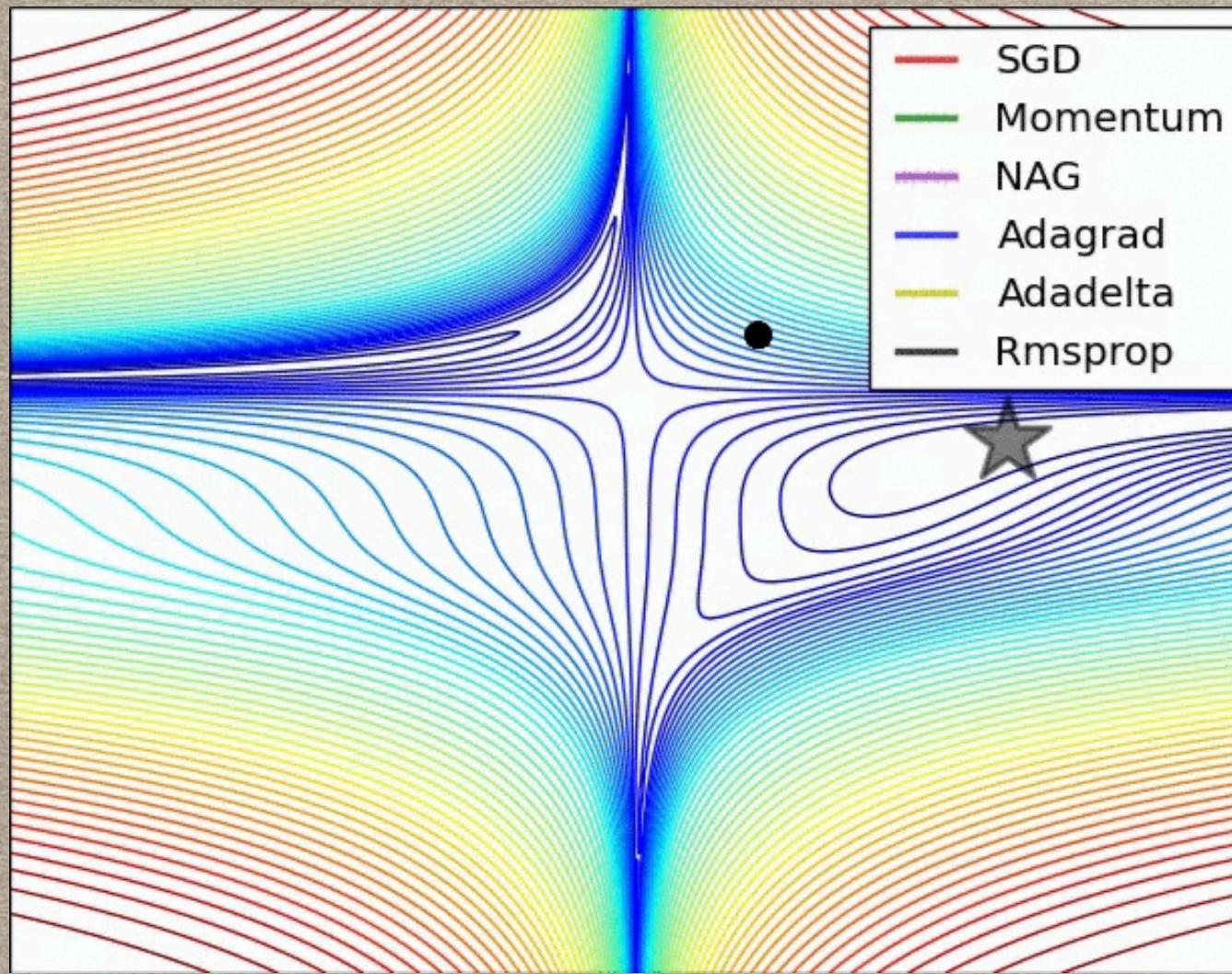
Optimisation functions usually calculate the gradient i.e. the partial derivative of loss function with respect to weights, and the weights are modified in the opposite direction of the calculated gradient. This cycle is repeated until we reach the minima of loss function.



The procedure of repeatedly evaluating the gradient and then performing a parameter update is called Gradient Descent.

IMPORTANCE OF THE LEARNING RATE

- * Learning rate will directly impact the training
- * Higher the rate faster is the decay of the algorithm
- * If you keep low learning rates, linear improvements will be seen
- * Advisable to have higher learning rates and then decrease it



QUESTIONS PLEASE!

github.com/vverdhan



Thanks

