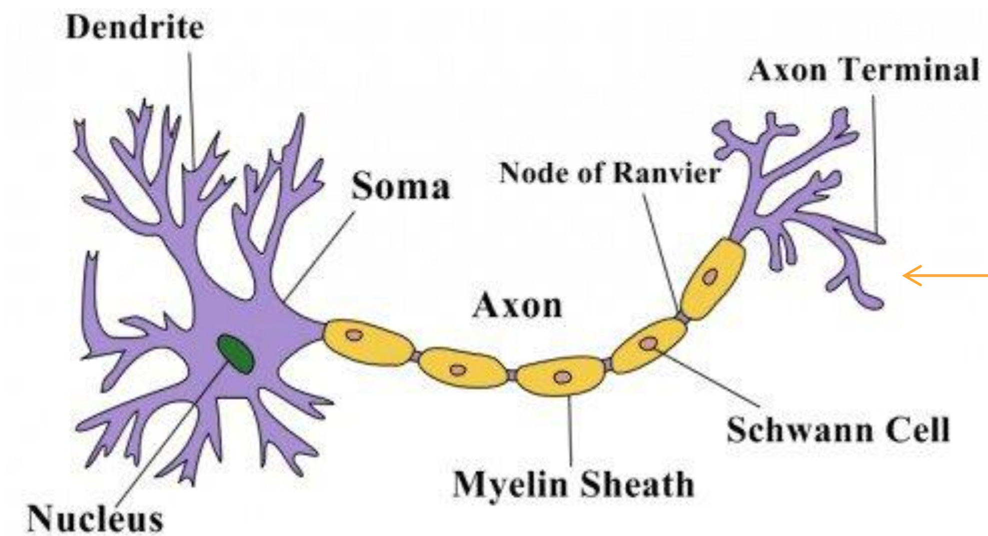
Neural Networks

Prof. Ankur Sinha

Indian Institute of Management Ahmedabad

Gujarat India

A typical Neuron

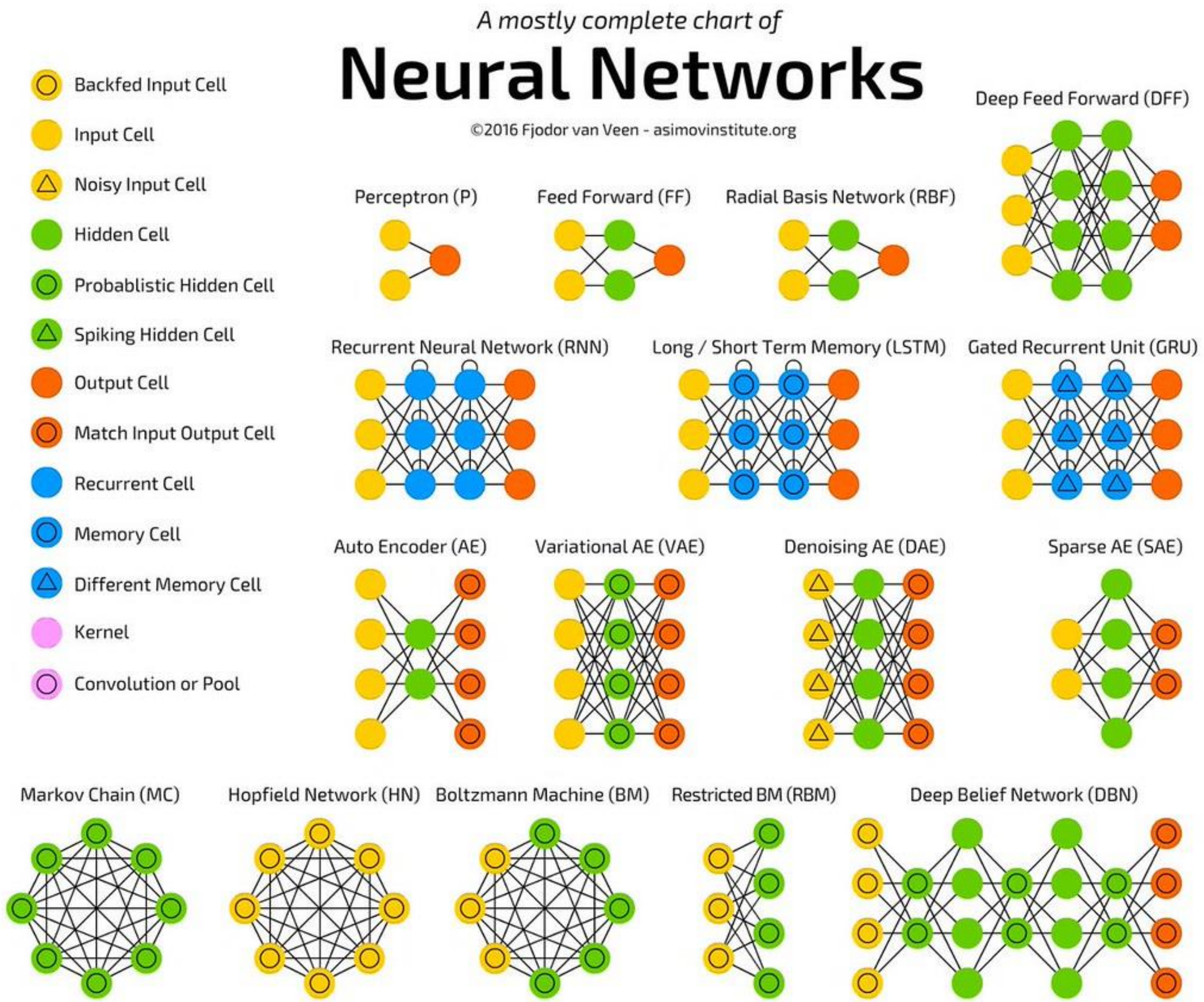


Releases neurotransmitters to other neurons

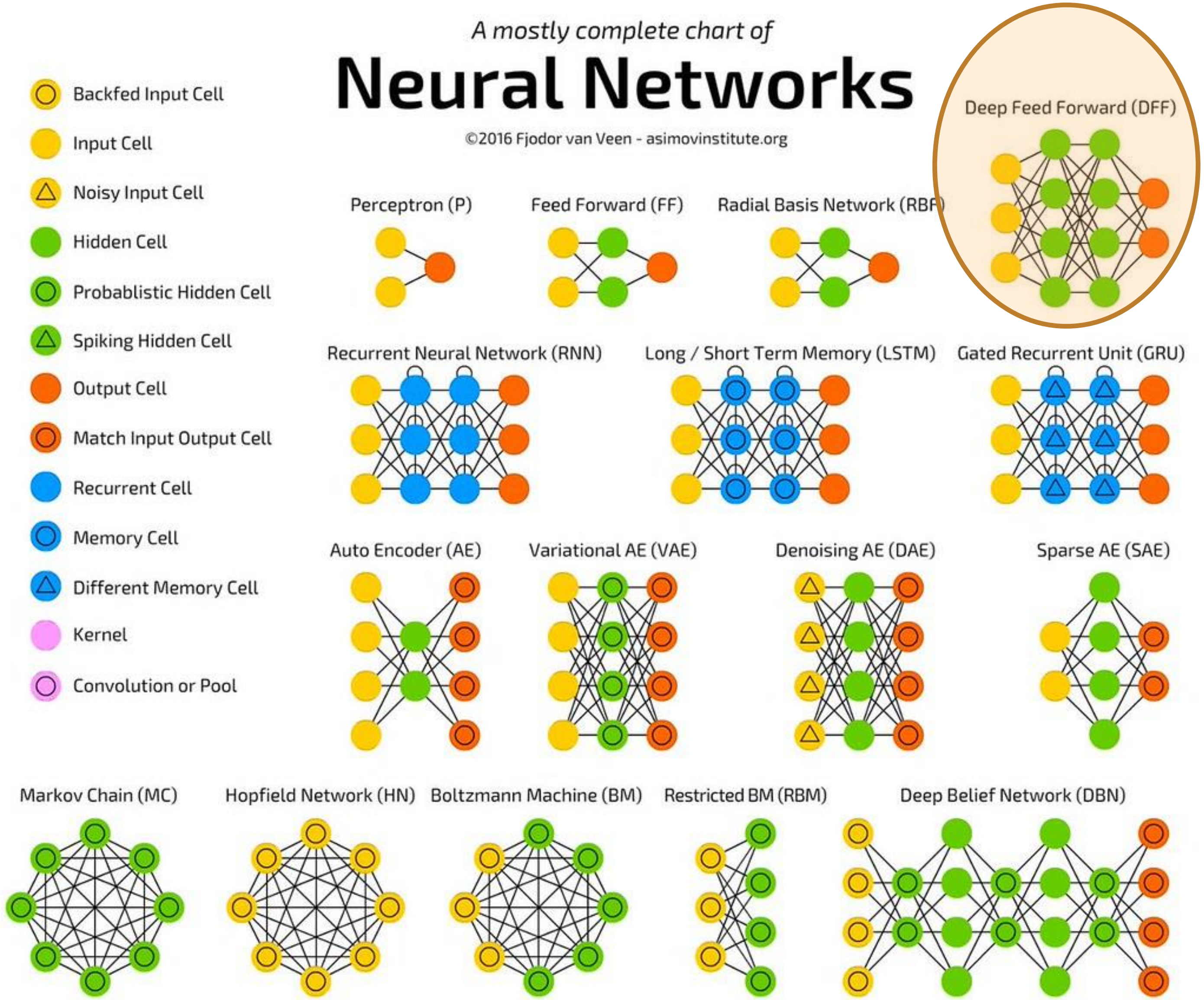
Information Flow

Applications

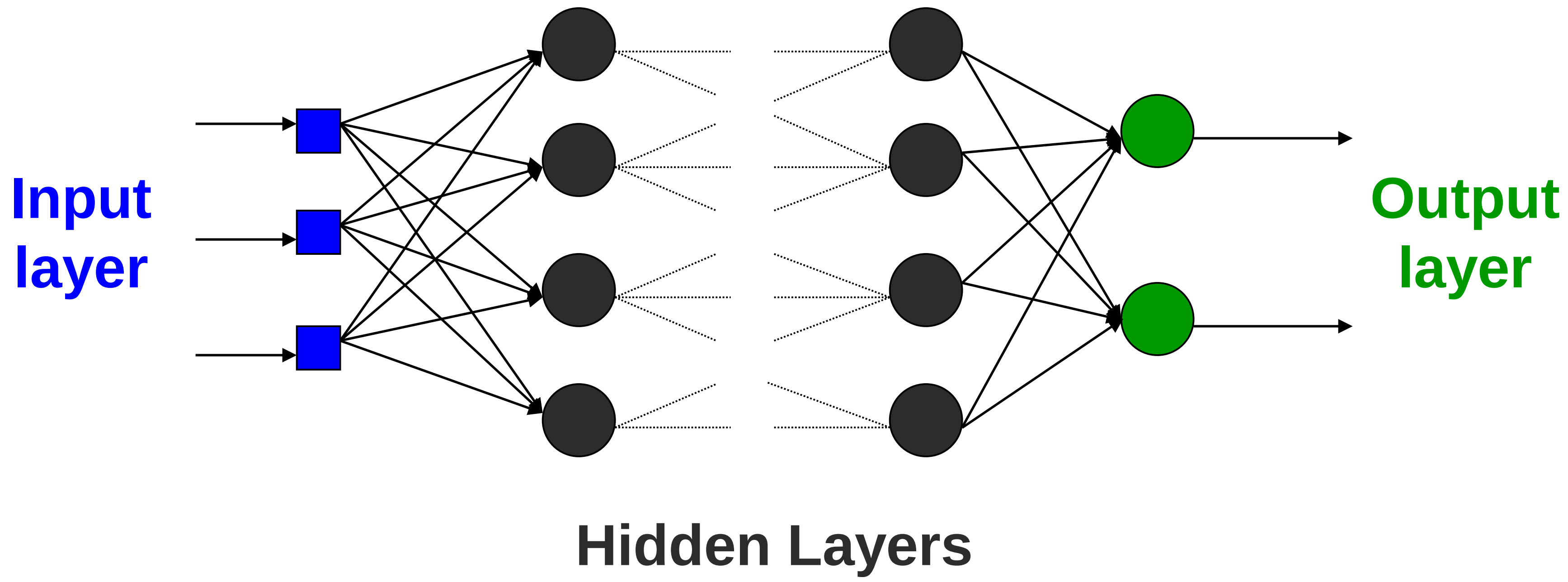
- Speech recognition
- Handwriting recognition
- Driverless Cars
- Products: Google translate, Alexa



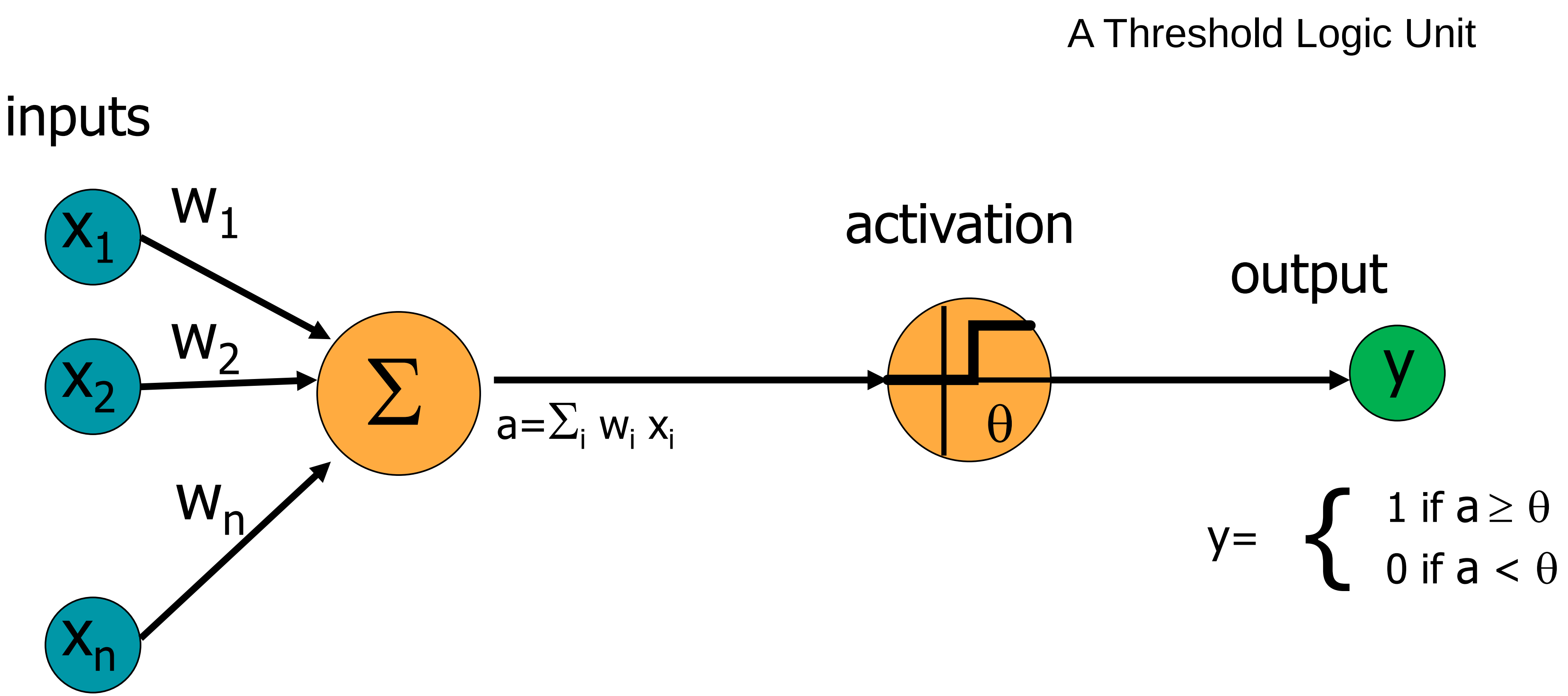
A mostly complete chart of
Neural Networks
©2016 Fjodor van Veen - asimovinstitute.org
Backfed Input Cell
Input Cell
Noisy Input Cell
Hidden Cell
Probablistic Hidden Cell
Spiking Hidden Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Different Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)
Markov Chain (MC)
Hopfield Network (HN)
Boltzmann Machine (BM)
Restricted BM (RBM)
Deep Belief Network (DBN)


A mostly complete chart of
Neural Networks
©2016 Fjodor van Veen - asimovinstitute.org
Backfed Input Cell
Input Cell
Noisy Input Cell
Hidden Cell
Probablistic Hidden Cell
Spiking Hidden Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Different Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)
Markov Chain (MC)
Hopfield Network (HN)
Boltzmann Machine (BM)
Restricted BM (RBM)
Deep Belief Network (DBN)

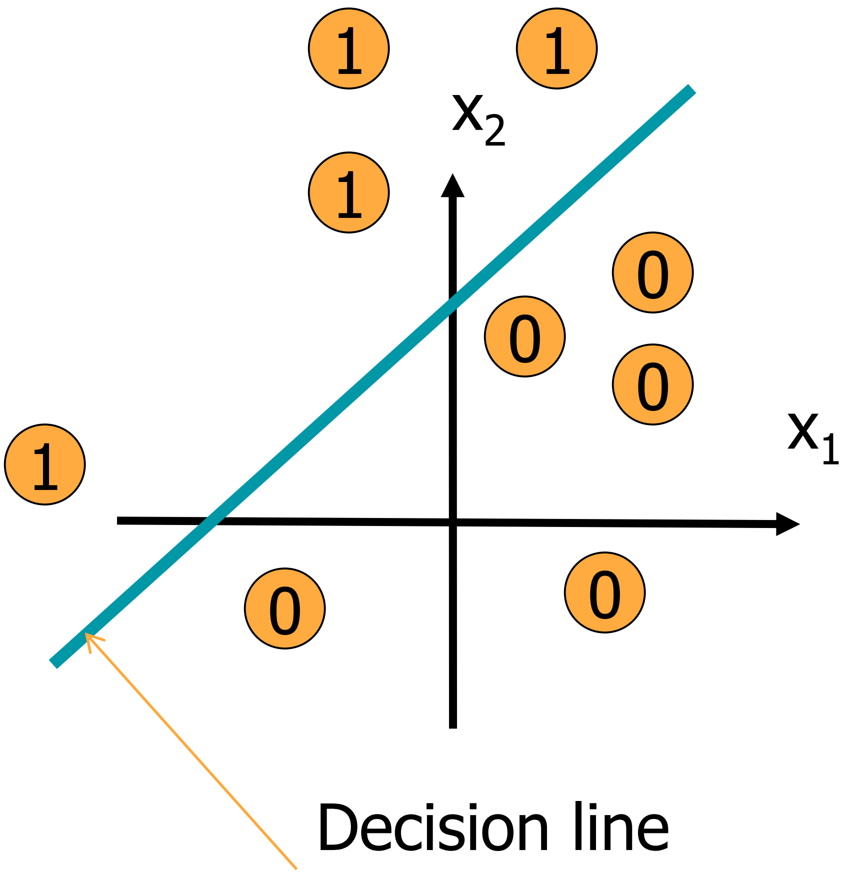
MLP Architecture



A Simple Architecture



Decision Surface of a TLU



Decision line

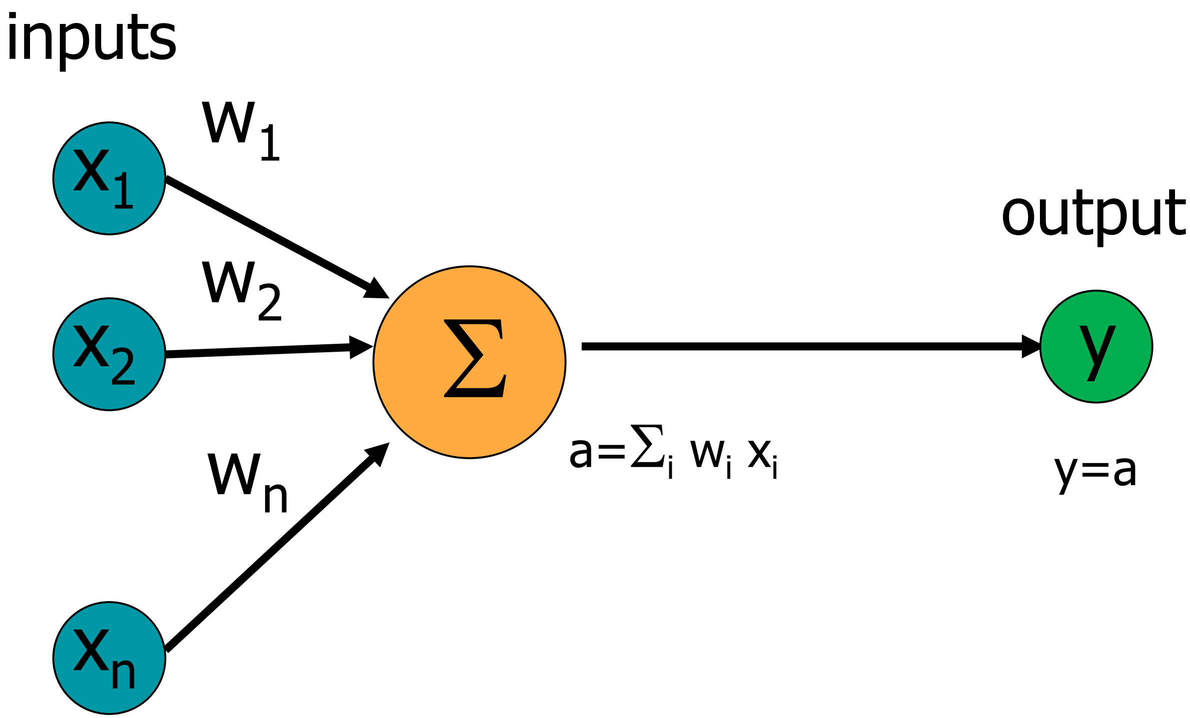
$w_1 x_1 + w_2 x_2 = \theta$

A TLU works as a linear classifier

Similar to SVM?

How do you identify the weights and threshold?

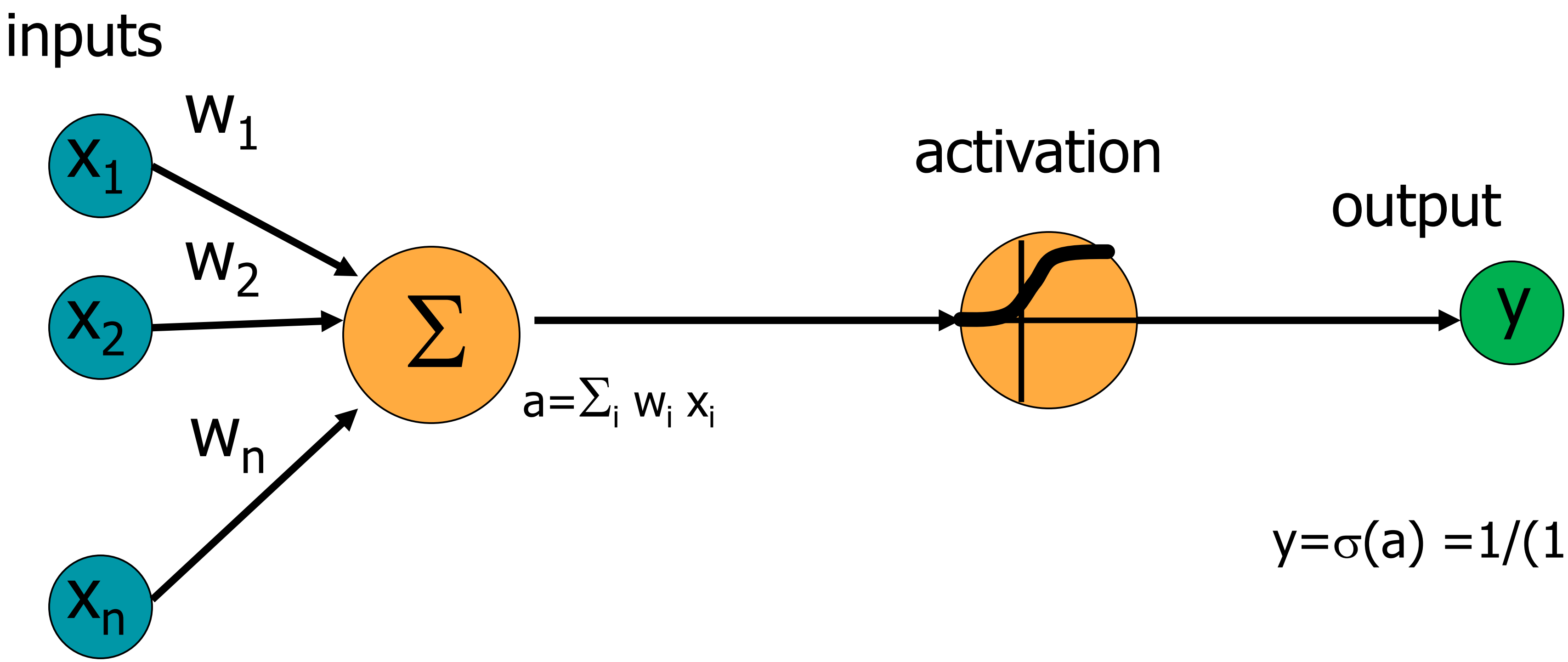
A Linear Unit



Tries to give the best linear relationship between input and output
Similar to regression?

Neuron with Sigmoid Function

A Threshold Logic Unit

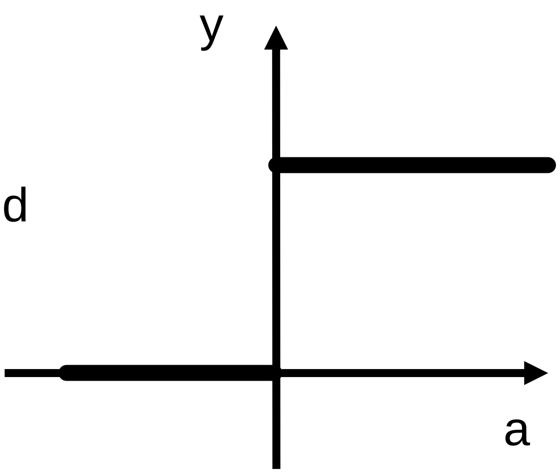


$y=\sigma(a) = 1/(1+e^{-a})$

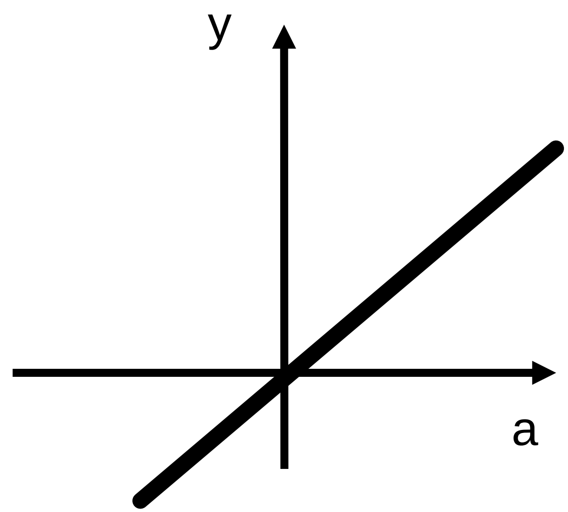
Gradient descent rules are used
to learn the parameters of the NN

Types of Activation Functions

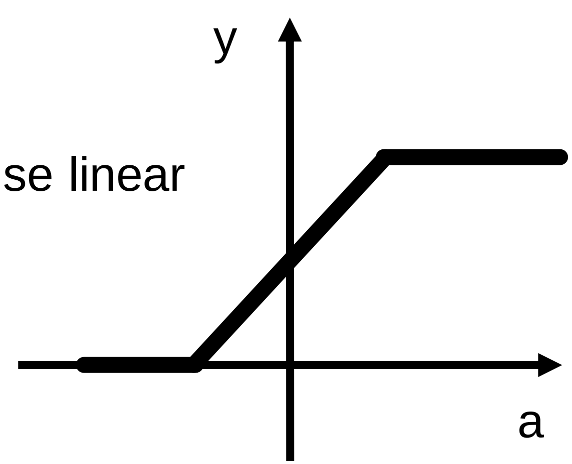
threshold



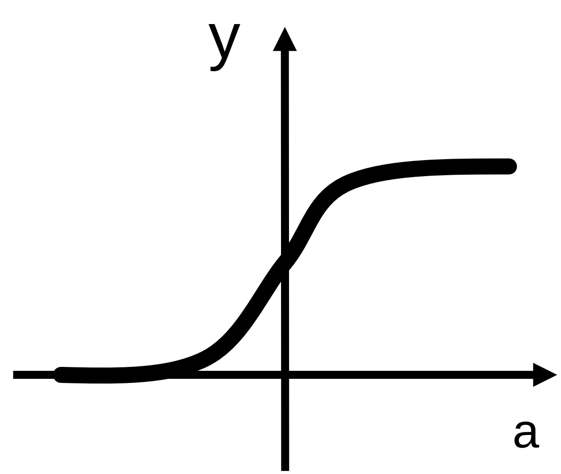
linear



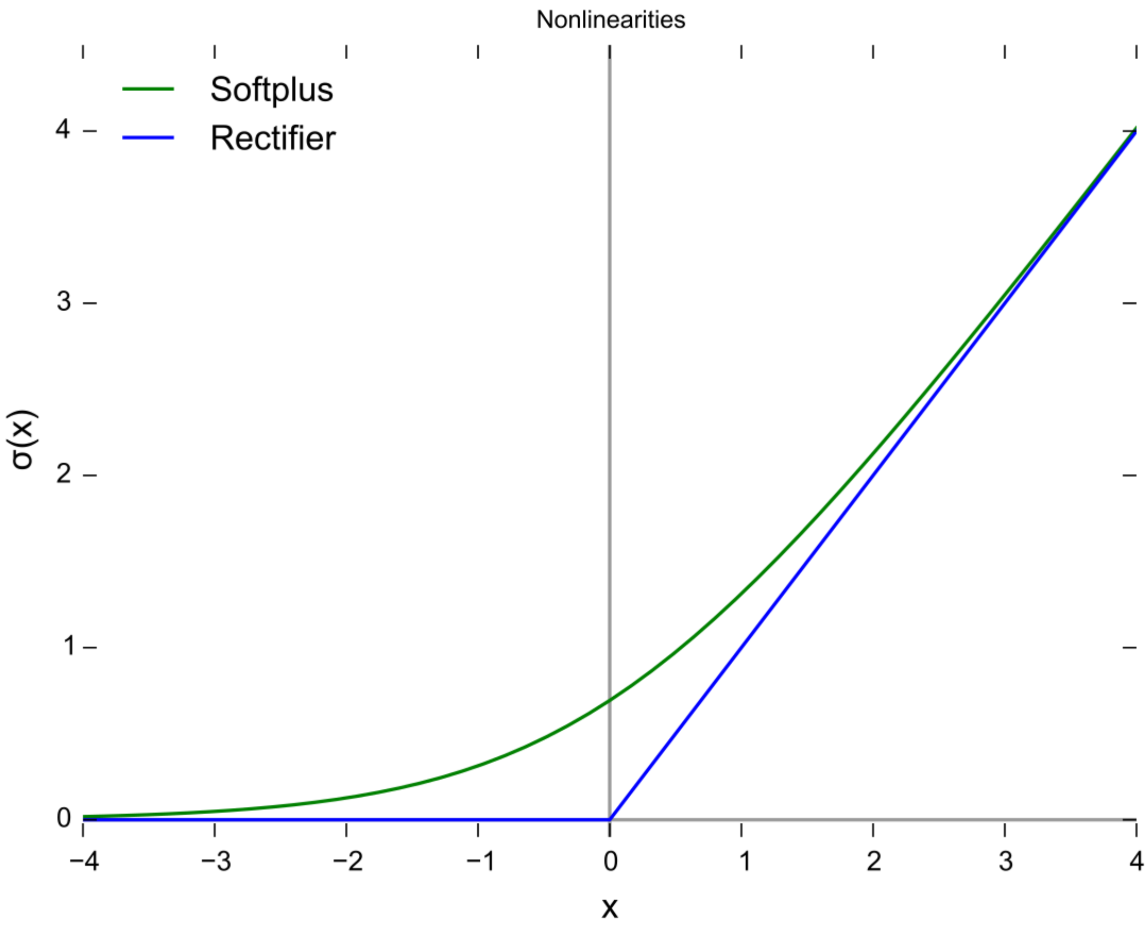
piece-wise linear



sigmoid



Types of Activation Functions



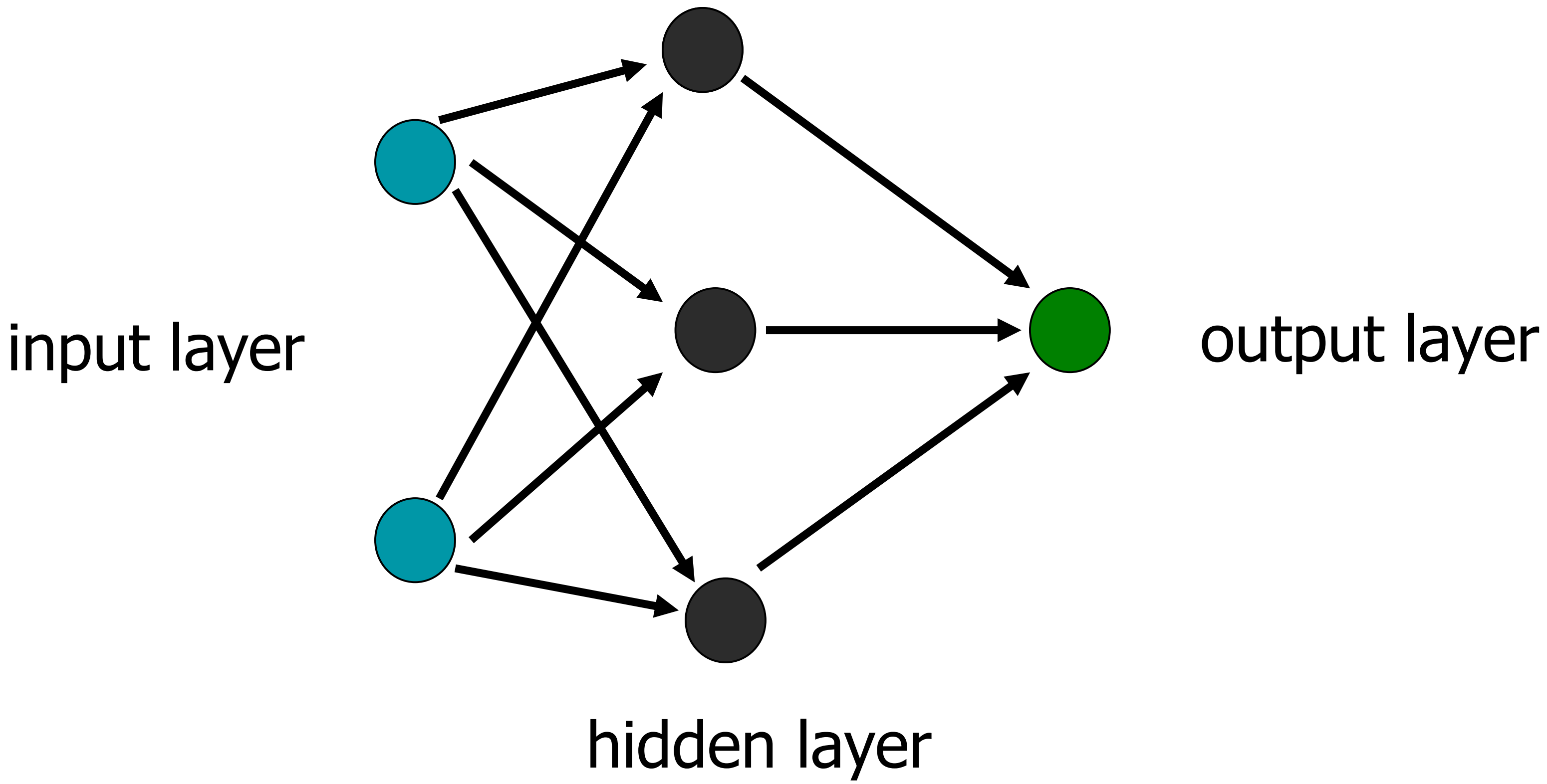
Training Neural Network

- A training set S of examples {**x**,t} is required
  - **x** is an input vector
  - t is the desired target vector
- Finding acceptable values of w and θ
  - Assume some values for w and θ
  - For the training example **x**, compute the network output y
  - Compare output y with targets t, a difference denotes error
  - Adjust w and θ so that the error can be reduced
  - Accept w and θ that leads to minimum error

Backpropagation Algorithm

- Refer to the separate set of slides

Multiple Layers



Backpropagation approach is used to train the neural network

More about NN Parameters

- The weights of the neural network are determined by training data
- As more training data is obtained the weights should be updated

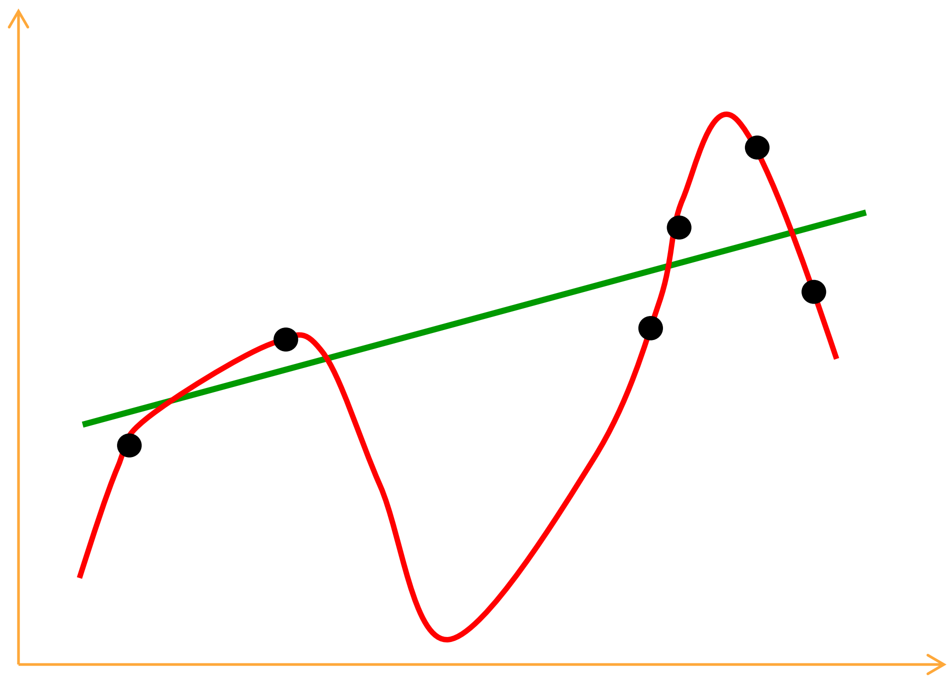
Neural Networks are Universal

- Any boolean function can be learnt by a neural network with single hidden layer
  - It might require a large number of hidden units

- Any mathematical function that is continuous and bounded can be approximated to an arbitrarily small accuracy using a neural network with one hidden layer
  - A large number of hidden units might be required if the error of approximation is very small

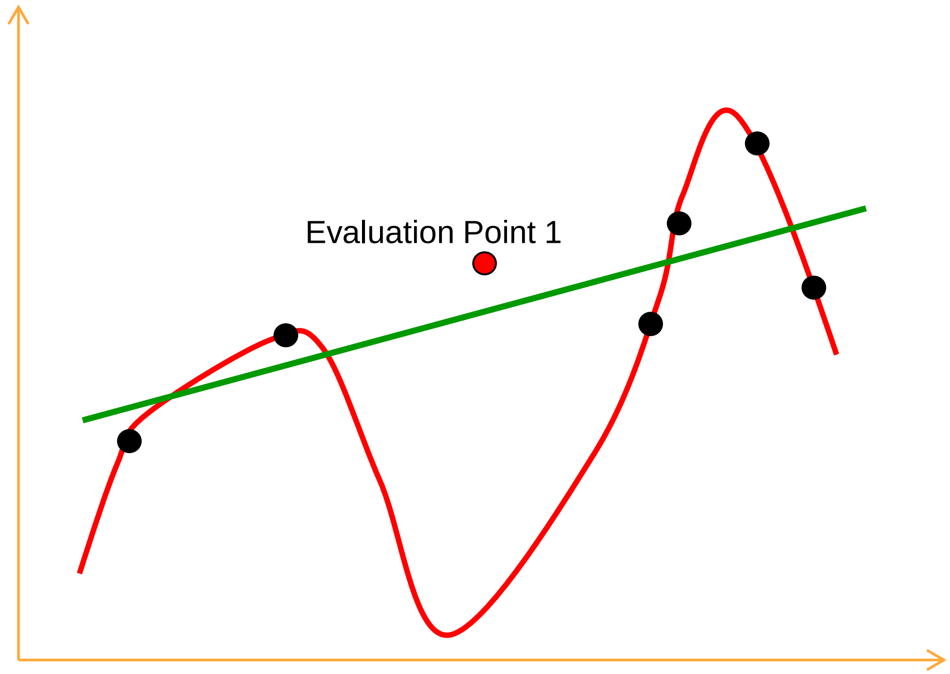
Be Careful!

- Neural network can easily lead to overfitting
- Try to minimize the generalization error than the training error



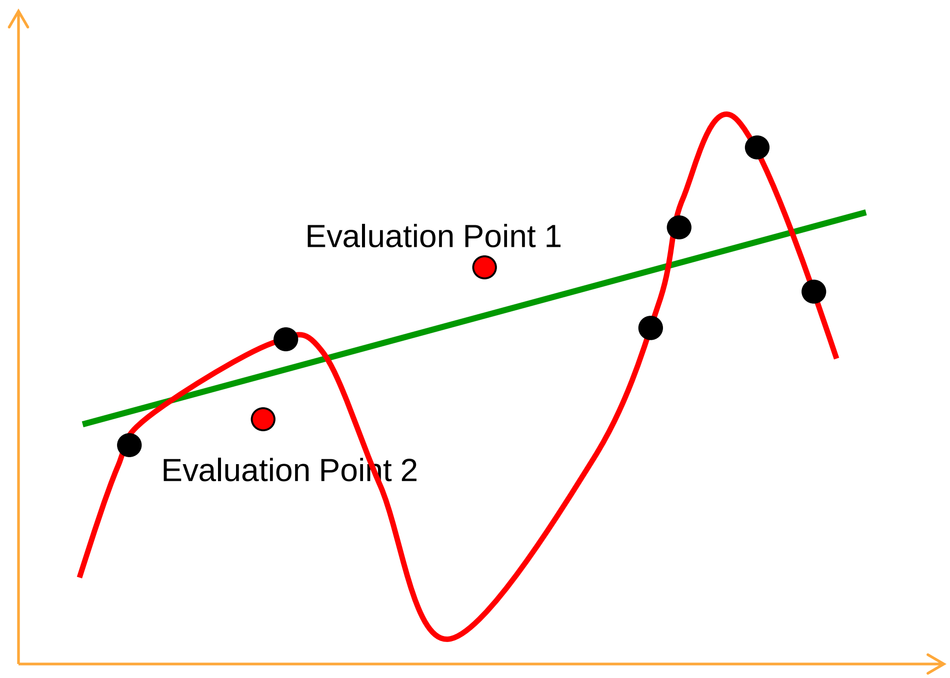
Be Careful!

- Neural network can easily lead to overfitting
- Try to minimize the generalization error than the training error



Be Careful!

- Neural network can easily lead to overfitting
- Try to minimize the generalization error than the training error



RNN, LSTM and CNN

- Recurrent Neural Networks (RNN)
  - It is a generalization of the feedforward network
  - It stores the output from the previous input and uses it along with the current input to produce the current output
  - Useful for connected tasks such as handwriting and speech recognition
- Long Short Term Memory (LSTM)
  - It is an extension of RNN and usually performs better as it has higher memory and resolves the vanishing gradient problem
  - Useful for classifying and predicting time series given time lags of unknown duration
- Convolutional Neural Network (CNN)
  - It is a feed forward neural network that uses filters and pooling
  - It is useful for handling images and spatial data, for instance, facial recognition, object detection, etc.

Text Data

- I liked the service
- It was horrible
- Waiting area was not so clean
- Wonderful experience

How do we vectorize the text?

Term Frequency

I liked the service

I	it	waiting	wonderful	liked	was	area	experience	the	horrible	not	so	clean	service
1	0	0	0	1	0	0	0	1	0	0	0	0	1

It was horrible

I	it	waiting	wonderful	liked	was	area	experience	the	horrible	not	so	clean	service
0	1	0	0	1	1	0	0	0	1	0	0	0	0

Waiting area was not so clean

I	it	waiting	wonderful	liked	was	area	experience	the	horrible	not	so	clean	service
0	0	1	0	0	1	1	0	0	0	1	1	1	0

Wonderful experience

I	it	waiting	wonderful	liked	was	area	experience	the	horrible	not	so	clean	service
0	0	0	1	0	0	0	1	0	0	0	0	0	0

Unigrams (information loss)

Does not account for sequence

I liked the service

I	it	waiting	wonderf ul	liked	was	area	experien ce	the	horrible	not	so	clean	servic e
1	0	0	0	1	0	0	0	1	0	0	0	0	1

It was horrible

I	it	waiting	wonderf ul	liked	was	area	experien ce	the	horrible	not	so	clean	servic e
0	1	0	0	0	1	0	0	0	1	0	0	0	0

Waiting area was not so clean

I	it	waiting	wonderf ul	liked	was	area	experien ce	the	horrible	not	so	clean	servic e
0	0	1	0	0	1	1	0	0	0	1	1	1	0

Wonderful experience

I	it	waiting	wonderf ul	liked	was	area	experien ce	the	horrible	not	so	clean	servic e
0	0	0	1	0	0	0	1	0	0	0	0	0	0

Unigrams (bigrams)

Some sequence is accounted

I liked the service

I	it	waiting	wonderful	liked	was	area	…	I liked	liked the	the service	was horrible	not so	so clean	…
1	0	0	0	1	0	0	…	1	1	1	0	0	0	…

It was horrible

I	it	waiting	wonderful	liked	was	area	…	I liked	liked the	the service	was horrible	not So	so clean	…
0	1	0	0	0	1	0	…	0	0	0	1	0	0	…

Waiting area was not so clean

I	it	waiting	wonderful	liked	was	area	…	I liked	liked the	the service	was horrible	not so	so clean	…
0	0	1	0	0	1	1	…	0	0	0	0	1	1	…

Wonderful experience

I	it	waiting	wonderful	liked	was	area	…	I liked	liked the	the service	was horrible	not so	so clean	…
0	0	0	1	0	0	0	…	0	0	0	0	0	0	…

How much importance to give to stop words?

- Words like the, of, on that, at are stop words that can be filtered out
- Filtering may lead to loss of information
- Can we do appropriate weighting?

Term frequency-Inverse Document Frequency

- Term-frequency is multiplied by a statistical weight called inverse document frequency
- Words that are present often in almost all text are often unimportant

Term frequency-Inverse Document Frequency

- Total number of documents/text: 1000,000
  - Term of interest:

a	present in 1000,000	$\mathrm{w} = \log \frac{1000{,}000}{1000{,}000} = 0$
person	present in 10,000	$\mathrm{w} = \log \frac{1000{,}000}{10{,}000} = 2$
personal	present in 1000	$\mathrm{w} = \log \frac{1000{,}000}{1000} = 3$
information	present in 100	$\mathrm{w} = \log \frac{1000{,}000}{100} = 4$

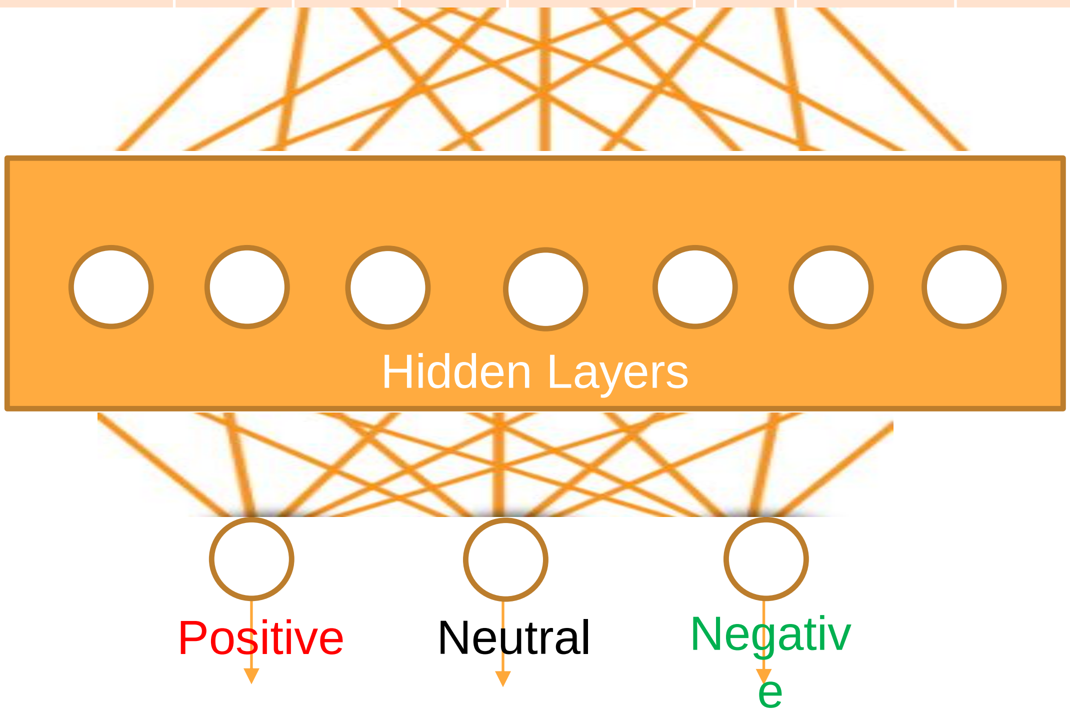
These weights can be a measure of importance. The term frequency is multiplied by these weights to get tf-idf vector.

I liked the service

Vectorize

I	it	waiting	wonderful	liked	was	area	experience	the	horrible	not	so	clean	service
1	0	0	0	1	0	0	0	1	0	0	0	0	1

Input



Output

Each of the output nodes fires a 0 or 1 (or the probability)

Sentiment Analysis in Finance

- Let us apply sentiment analysis to financial text

Dataset references:

Sinha, A., Kedas, S., Kumar, R., & Malo, P. (2022). SEntFiN 1.0: Entity-aware sentiment analysis for financial news. Journal of the Association for Information Science and Technology.

(https://www.kaggle.com/datasets/ankurzing/aspect-based-sentiment-analysis-for-financial-news)

Malo, P., Sinha, A., Korhonen, P., Wallenius, J., & Takala, P. (2014). Good debt or bad debt: Detecting semantic orientations in economic texts. Journal of the Association for Information Science and Technology, 65(4), 782-796.

(https://www.kaggle.com/datasets/ankurzing/sentiment-analysis-for-financial-news)