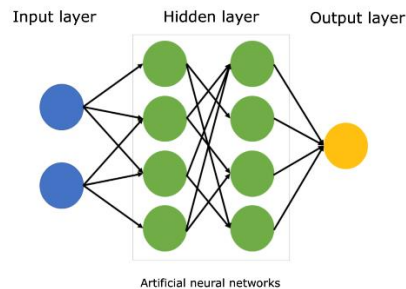# Artificial Neural Network (ANN)

ANN is a Technology that mimics a human brain to learn from some key features and classify or predict in the real world.


Artificial neural networks

1) Input Layer – To get the data from the user or a client or a server to analyze and give the result.
2) Hidden Layer – This layer can be in any number and these layers will analyze the inputs with passing through them with different biases, weights and activation functions to provide an output.
3) Output layer – This is where we can get result from a neural network .

Training a neural network means that we find the optimal values of the weights and the weights associated with the bias.

Recurrent Neural Networks (RNN)

When you give an input in RNN, the output is calculated, but at the same time output is passed again, the second feature of the input is given.
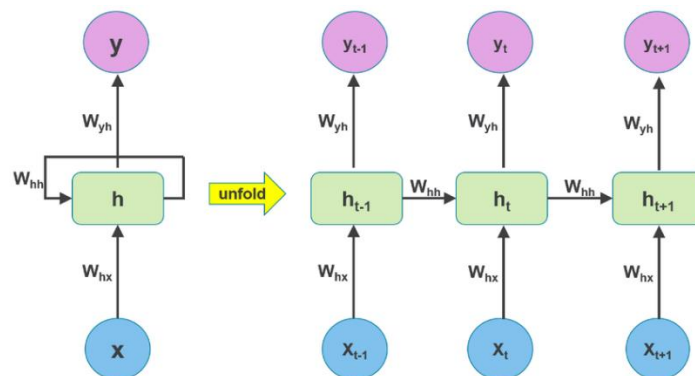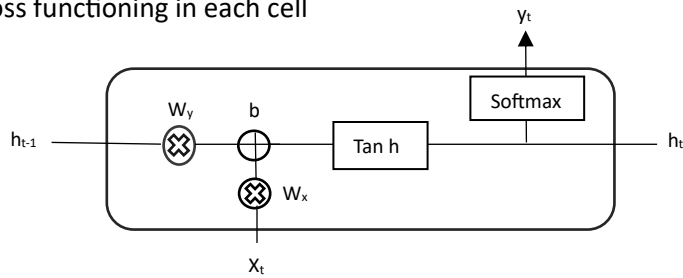

In NN you give all the features at the same time, but in RNN you give all the features one step at a time.

In the above diagram , each of the nodes except the first one you have two inputs . One of them is the output of the previous time step and one of them is the input of this time step.

Mathematical Calculation

$y_t = a(W_x x_t + W_y y_{t-1} + b)$

Depicting the cross functioning in each cell

This is the cell in which you the output to the outer world as well as to the next cell.

Types of RNNs

1) Sequence to Sequence – price forecasting
2) Sequence to Vector – scam or not scam
3) Vector to Sequence – image captioning

Shortcomings of simple RNN cells

- Unstable gradients : use techniques for unstable gradients or layer normalization.
- The long term memory does not work well. forgets the beginning of long sentences .

Solution to the problem : Use LSTM or GRU cells.

LSTM (Long Short Term Memory Memory) cell
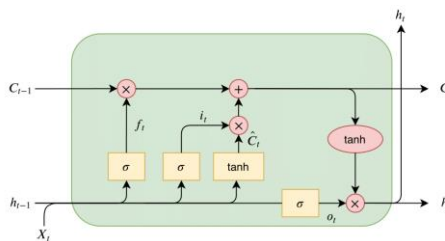
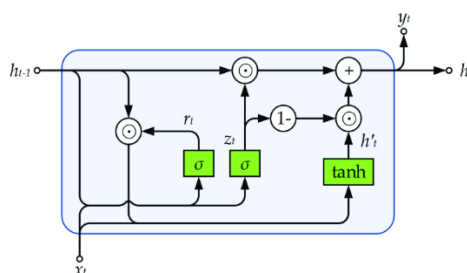σ – sigmoid function

tan h – tangent hyperbolic function

$C_{(t-1)}$, $C_{(t)}$ – long term memory hidden state.

- We forget or add things to this hidden state

Gates – create info workflow

- How info is used in this specific time.

GRU(Gated Recurrent Unit)

-> Simplified version of an LSTM cell

-> There is no separate output, the whole state is output at evey step.

-> The gate controller 'r' decides which part of the previous state will be shown to the main layer 'y'.

Transformers

-> Type of neural network.

-> BERT, GPT-3, T5 are based on transformers.

What are Transformers?

->Basically, they were invented to solve the problem of low memory storage faced during using RNN.

Problem with RNN:-

-> example – Let's say, you want to translate a statement from English to French .

-> an RNN would take an English sentence as input process the words one at a time and then sequentially spit out there French counterparts.

-> While translating a statement the sequence of the words is very important.

-> On shuffling the words the words of the sentence the meaning of it totally changes.

-> But, RNN (Recurrent Neural Networks) always had a problem while handling large sequences of text like long paragraphs or essays.

-> By the time, the end of a paragraph is analyzing they'd forget the beginning of it.

-> Also, RNN were pretty hard to train. Because they process words sequentially they couldn't parallelize well.

Solution:-

-> Transformers were developed in 2017 by google researchers and the university of Toronto. And they were initially designed for translation, but transformers can easily be parallelized.

-> They meant with right hardware, you can train really big modules.

How Transformers work?

 3 innovations making transformers work so well:-

1) Positional Encoding
2) Attention
3) Self-Attention

Examples of Transformers:-

GPT – Generative Pre-Trained Transformers

BERT – Bidirectional Encoder Representation from Transformers

- Used in google search to understand search queries
- Used in text summarization, question answering , classification and finding similar sentences.

T-5 – Text to Text Transfer Transformer

- Chatbots, Machine Translation system, Text summarization tools, code generation and robotics .

- Developed By Google in 2019