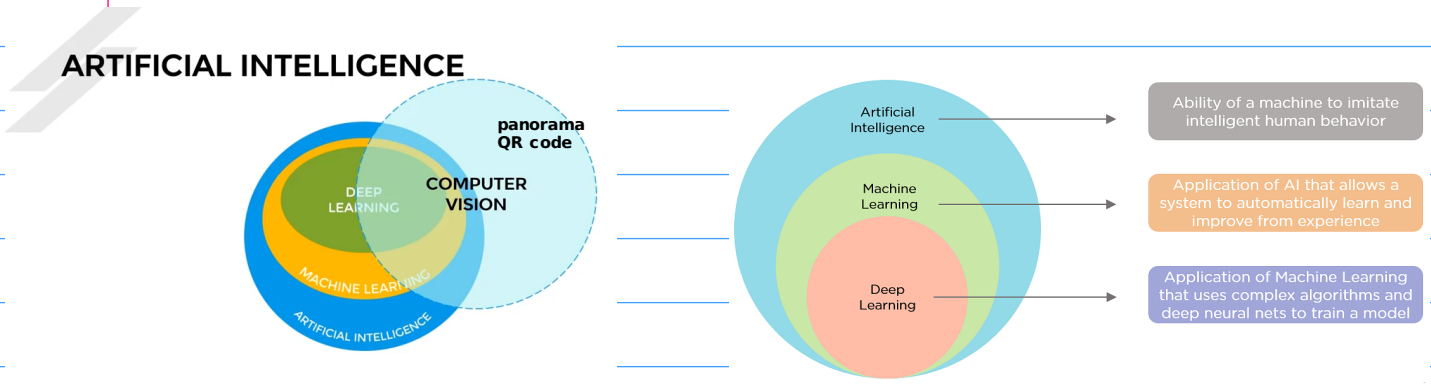
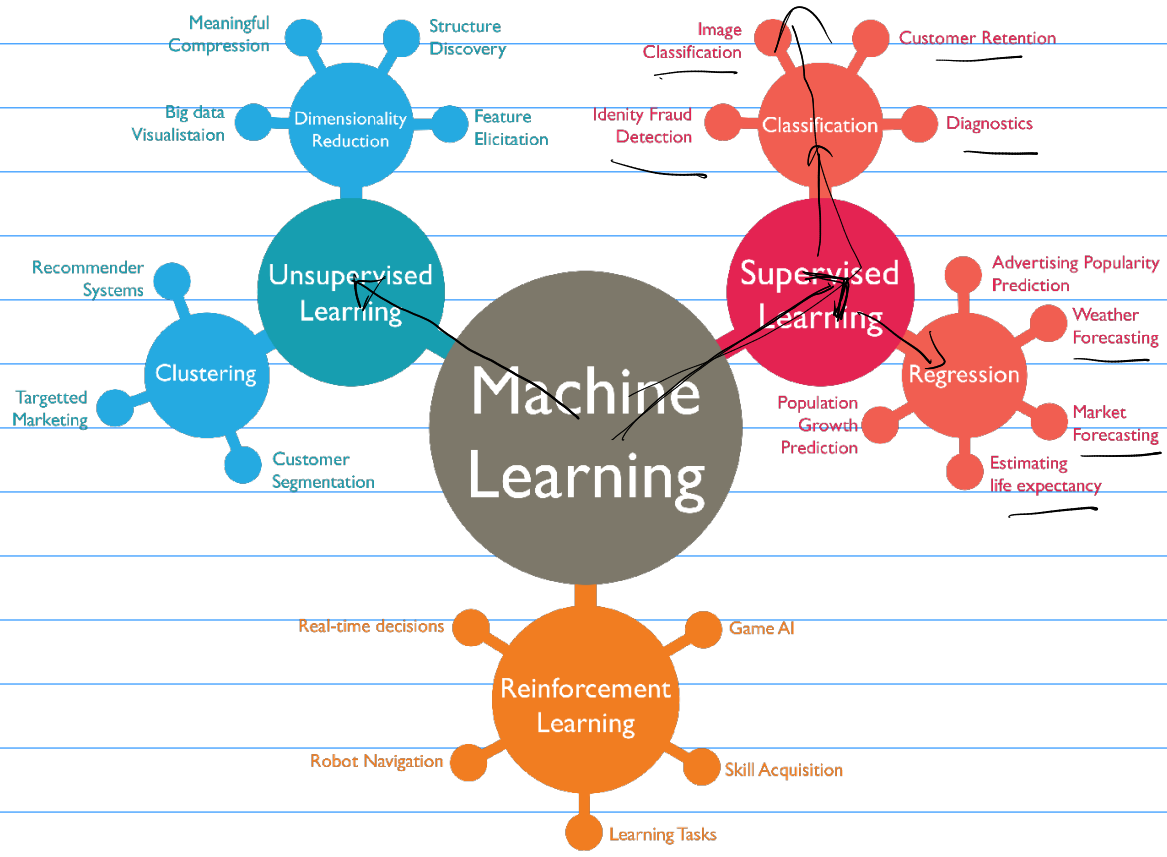
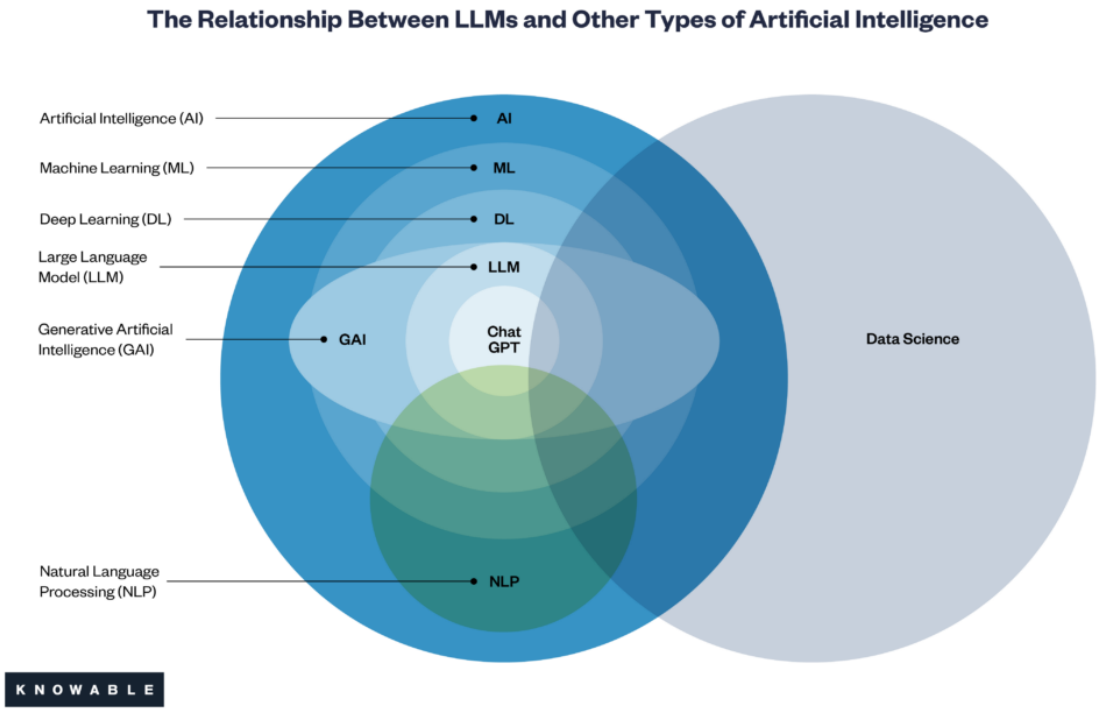
1. Introduction

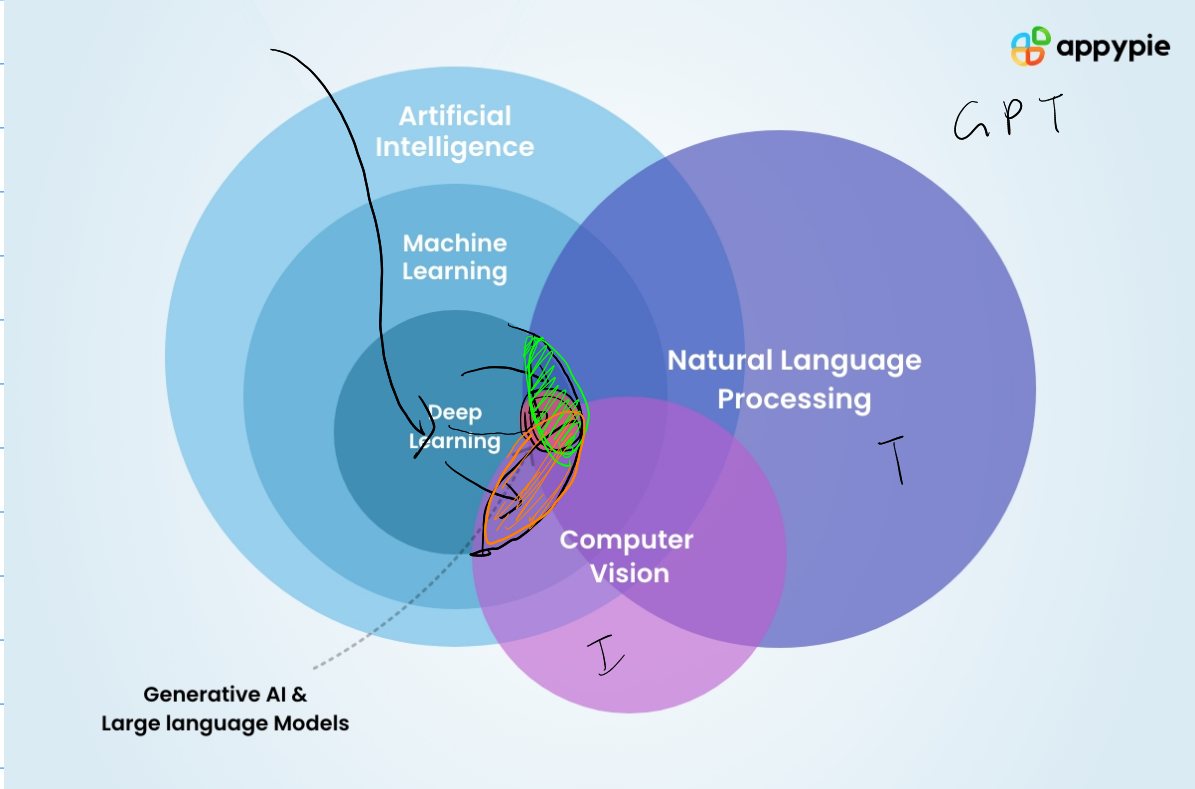
1.1 Overview



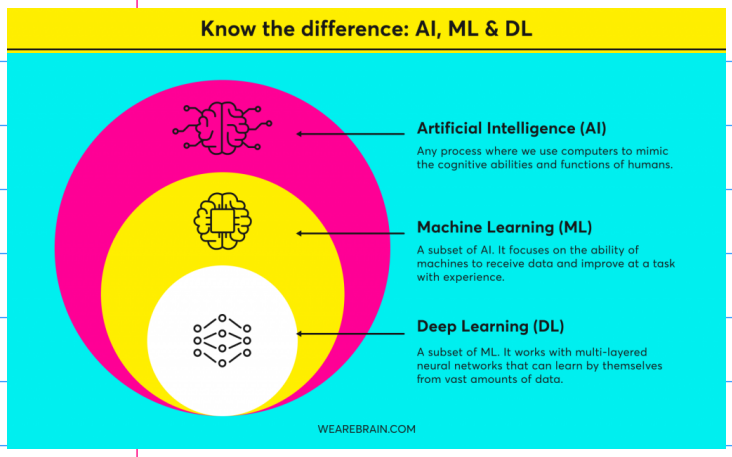
1.2 Machine Learning Types

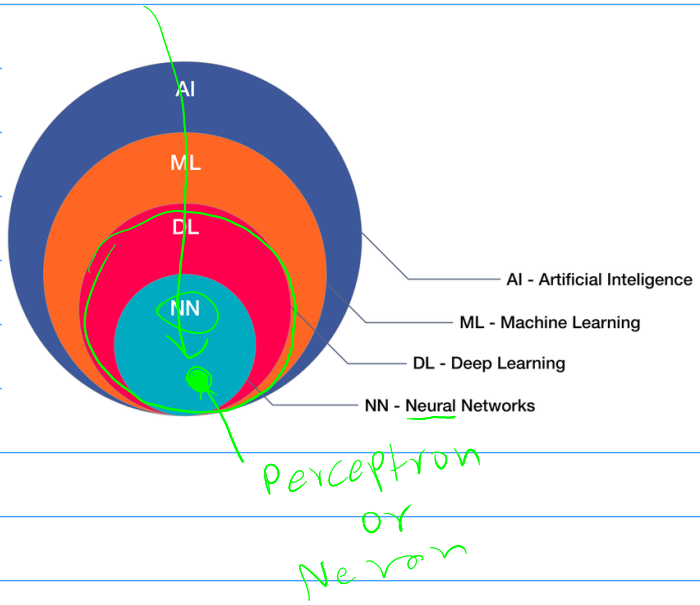



The Relationship Between LLMs and Other Types of Artificial Intelligence
Artificial Intelligence (AI)
AI
Machine Learning (ML)
ML
Deep Learning (DL)
DL
Large Language Model (LLM)
LLM
Generative Artificial Intelligence (GAI)
GAI
Chat GPT
Data Science
Natural Language Processing (NLP)
NLP
KNOWABLE

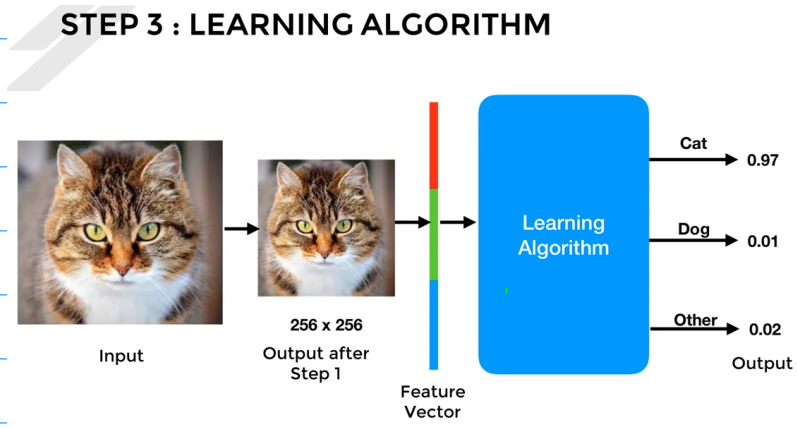

appypie
GPT
Artificial Intelligence
Machine Learning
Deep Learning
Natural Language Processing
Computer Vision
Generative AI & Large language Models

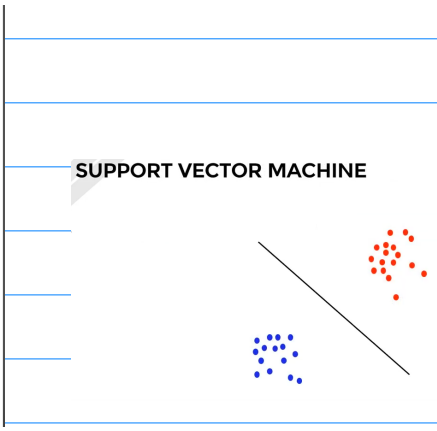
1.3 Deep Learning (DL)



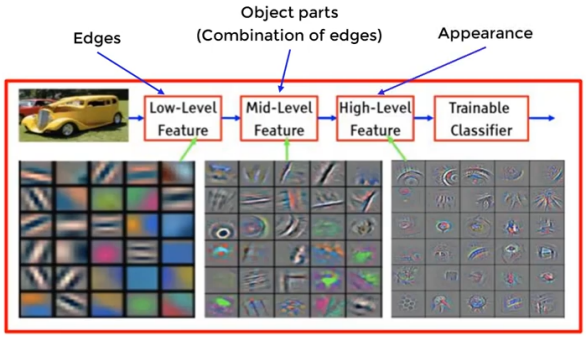


1.3 Why is Deep Learning so popular? ->





1. Learns Multi-level Representations

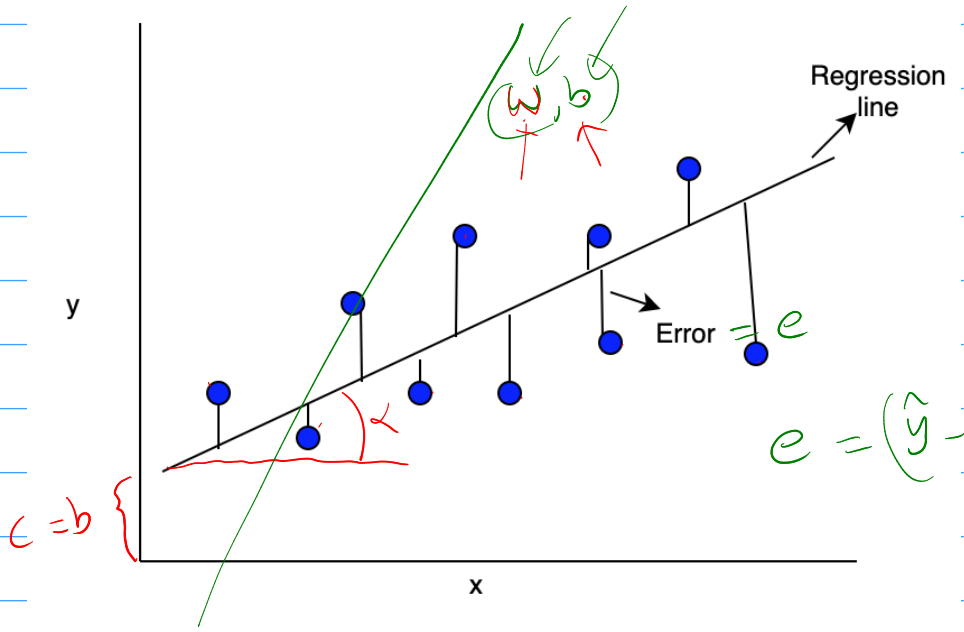


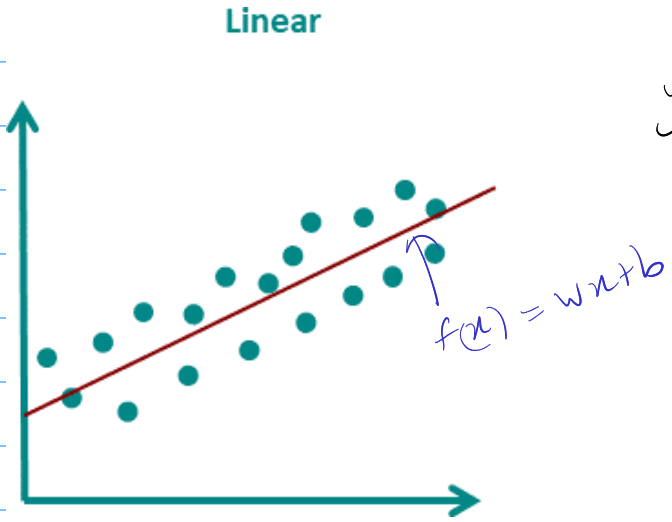
Learning has become so popular

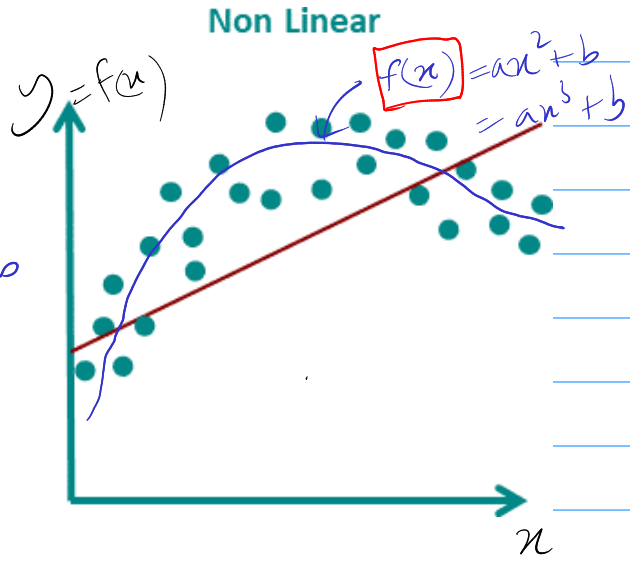
Linear Regression

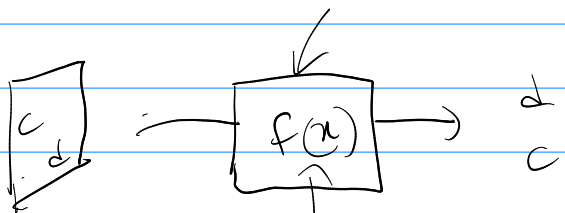
$y = mx + c$ ← $m = w = \tan\alpha$

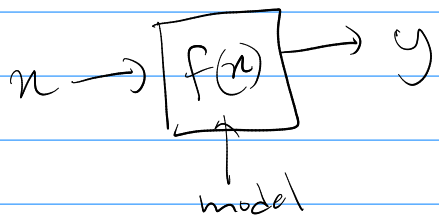
$y = wx + b$ $\quad c = b$

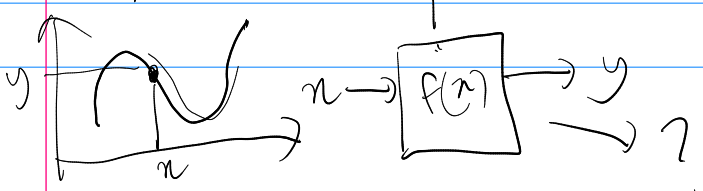




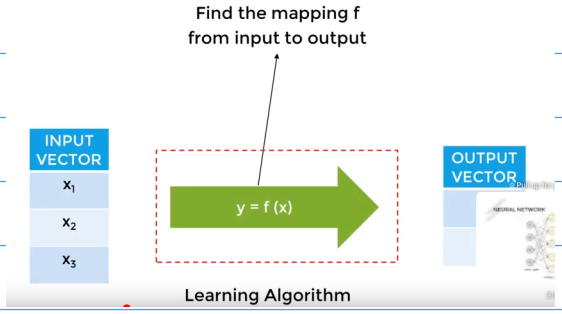


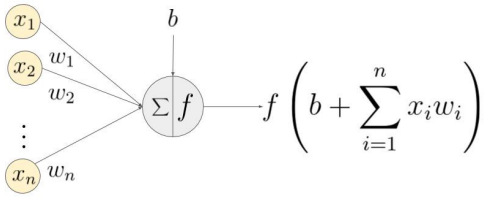




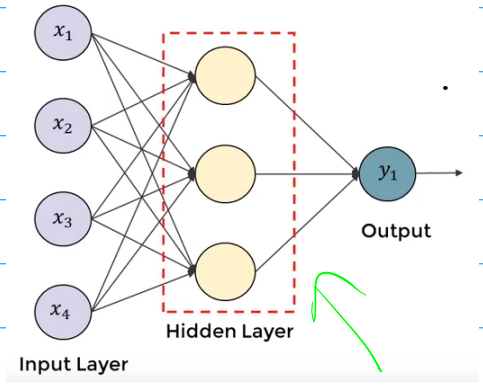


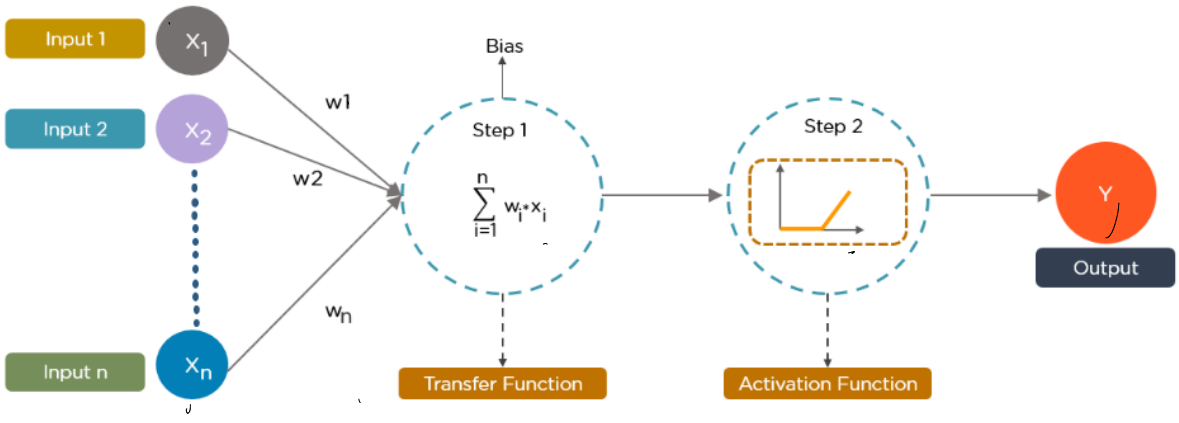
1.6 Neural Networks



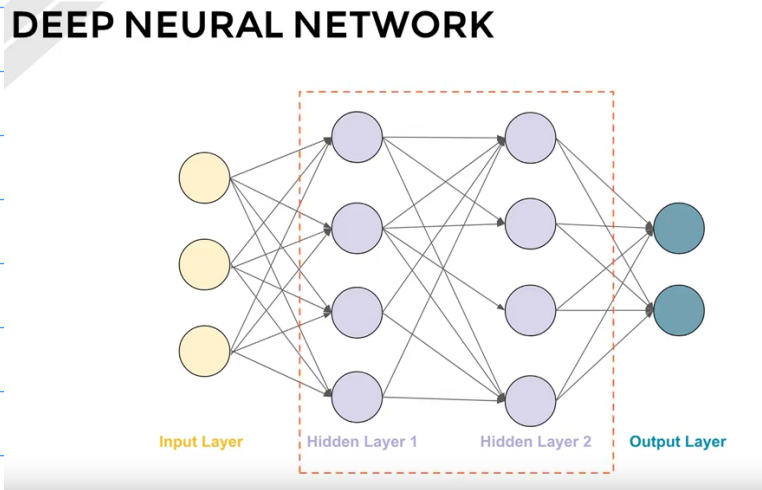


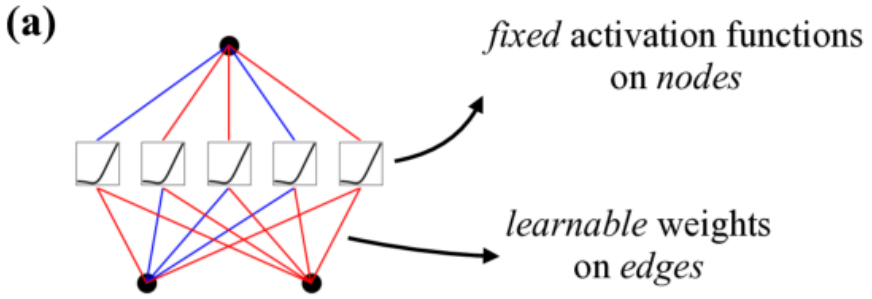
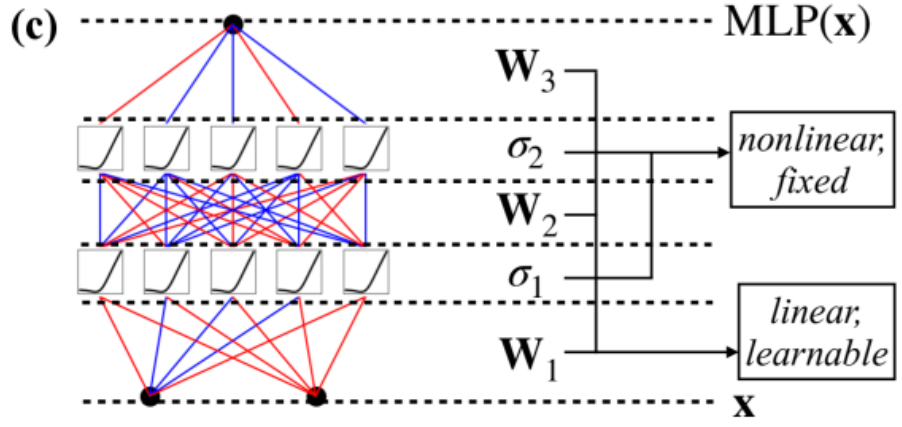
An example of a neuron showing the input (x_1 - x_n), their corresponding weights (w_1 - w_n), a bias (b) and the activation function ƒ applied to the weighted sum of the inputs.



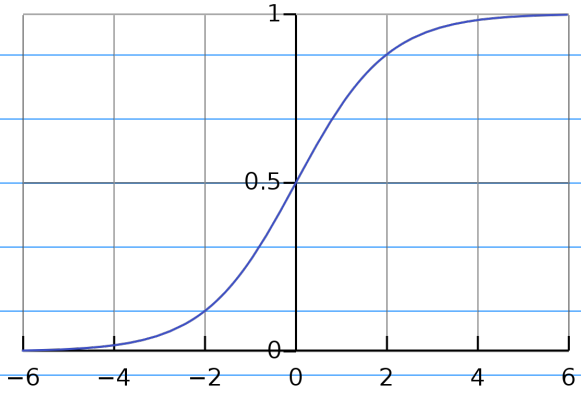


1.7 Deep Neural Networks



Model	**Multi-Layer Perceptron (MLP)**
Theorem	**Universal Approximation Theorem**
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(\epsilon)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$
Model (Shallow)	
Formula (Deep)	$\text{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$
Model (Deep)	

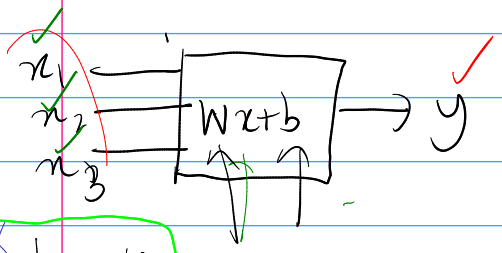
$$\sigma(x) = \frac{1}{1+e^{-x}}$$



$$y = wx + b$$

$$w = \begin{bmatrix} & & \end{bmatrix} \qquad w = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \end{bmatrix}$$



$$y = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + b$$

$$y = mx + b$$

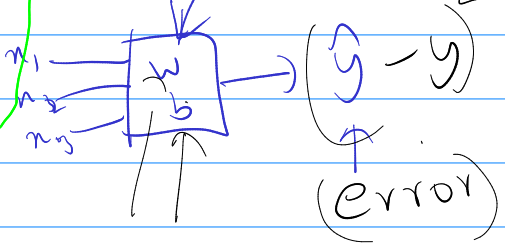
$$y = w_1x_1 + w_2x_2 + w_3x_3 + b$$

$$y = \sum_{i=1}^{n} w_i x_i + b$$

Linear

w_s ✓
b_s ✓

parameters



90%

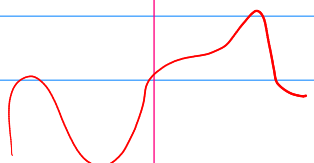
$y = mx + b$ →

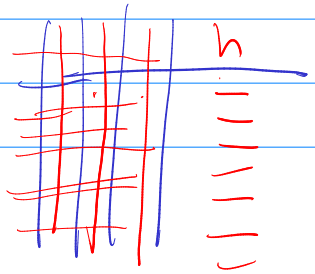
$$y = \sigma\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

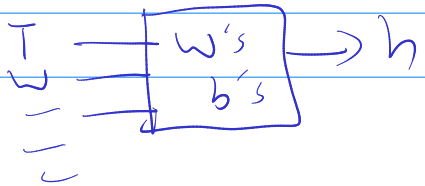
w's b's

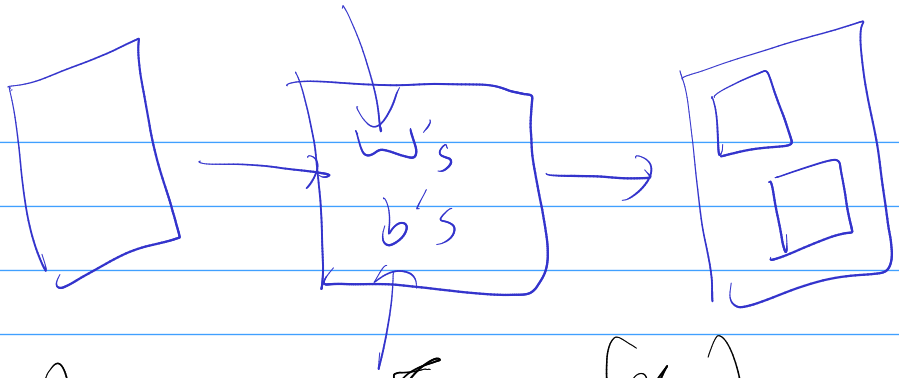
Activation function

σ









$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sigma\left(W_{n \times n} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} + b_{n \times 1} \right)$$

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ & & \\ w_{n1} & & w_{nn} \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Parameter → 7B, 30B