Day 2 (4 oct, 2024)

→ Neural Networks are superior than the traditional linear Regression or the linearity function

→ Non-linearity function such as activation ction (tanh, Sigmoid) are used in Neural Networks.

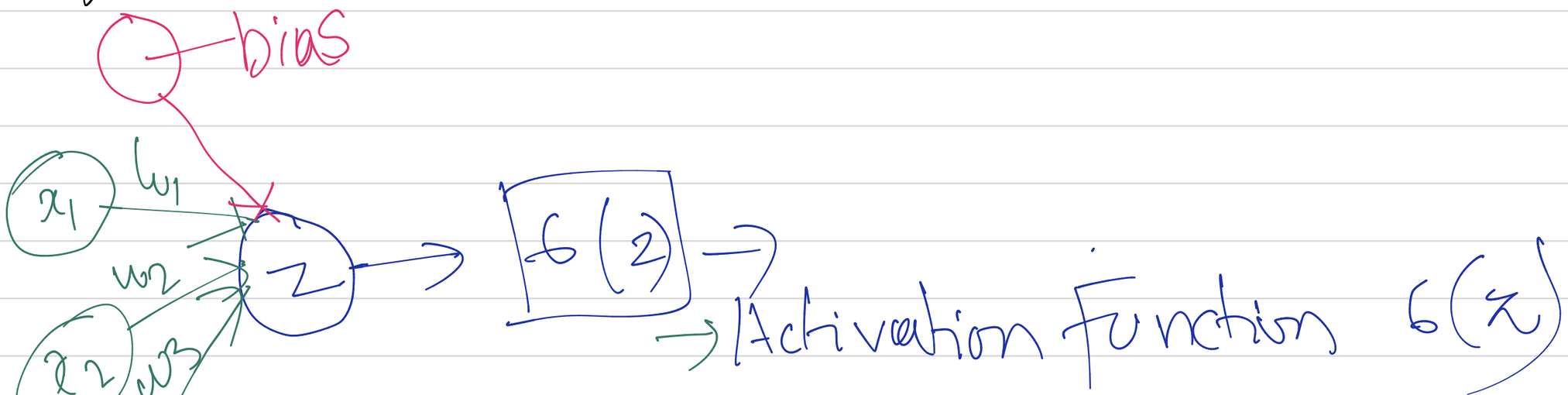
Sigmoid function with linear function

We have,

$$z = w_n x_n + bias$$

So, if we sketch the figure with single neuron

a neuron is a basic element or single node of Neural network



$$\sigma(z) \Rightarrow \frac{1}{1+e^{-z}}$$

$$A = \sigma(wx+b)$$ then

$$A \Rightarrow \frac{1}{1+e^{-(wx+b)}}$$

→ Applying Activation function can help to model and understand the complex relationships between Input and Output.

→ for Different Models we can or have to use various Suitable Activation functions.

→ these are different methods to connect 'w' weight or 'b' to the input and the neuron

and the Neural network. Like in CNN, RNN

→ In Image, we have structure of Image as matrix of pixels arrangement

→ In RNN we have the sequence of words, or sequence of numbers

So the arrangements are different and we call it inductive bias and enforce into the parameter 'w' and process with the activation function itself.

- Our parameters are feed into the model such that our model can distinguish and process the parameters.
  → See more details on Graph Neural Network

Represent Object as Data points

1) Human Eye and Brain can infer and Analyze the structures and shape.

→ But for Computer, it is Matrix or the numbers or pixels of captured Image objects -

Just difference in Single Pixel or 1 pixel in a Image can make our Computer difficult to understand.

→ Concept from Image Processing

→ So, Complexity in the analysis of object

→ Structure makes Computer vision or the Computer hard to understand the object's shape and data points.

Image as a function of Spatial Operation

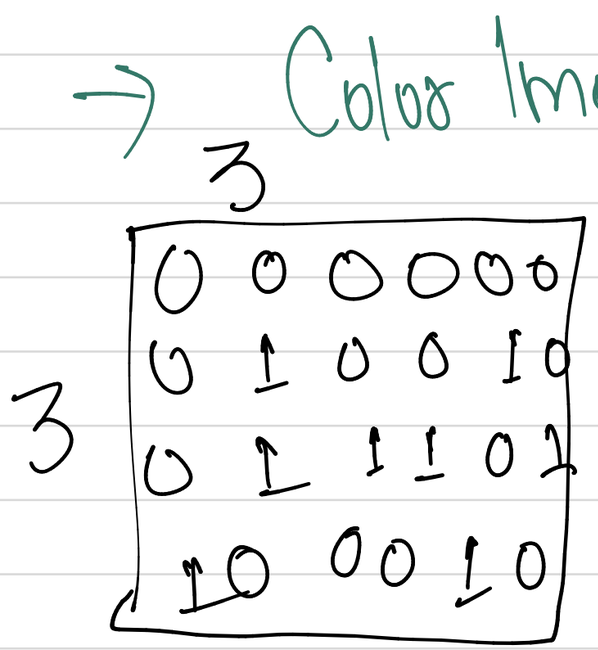
→ Images are 2D objects, Computer just see the numbers

→ 2D Image, each discrete point or location $f(x,y)$ consist of intensity value

→ 3D Image of discrete location $f(x, y, z)$
with the intensity level

→ Grayscale Image Represented as $M \times N$ or $N \times M$
matrix with values range (0 - 255 or 0/1)

→ Color Image RGB $3 \times N \times M$ matrix

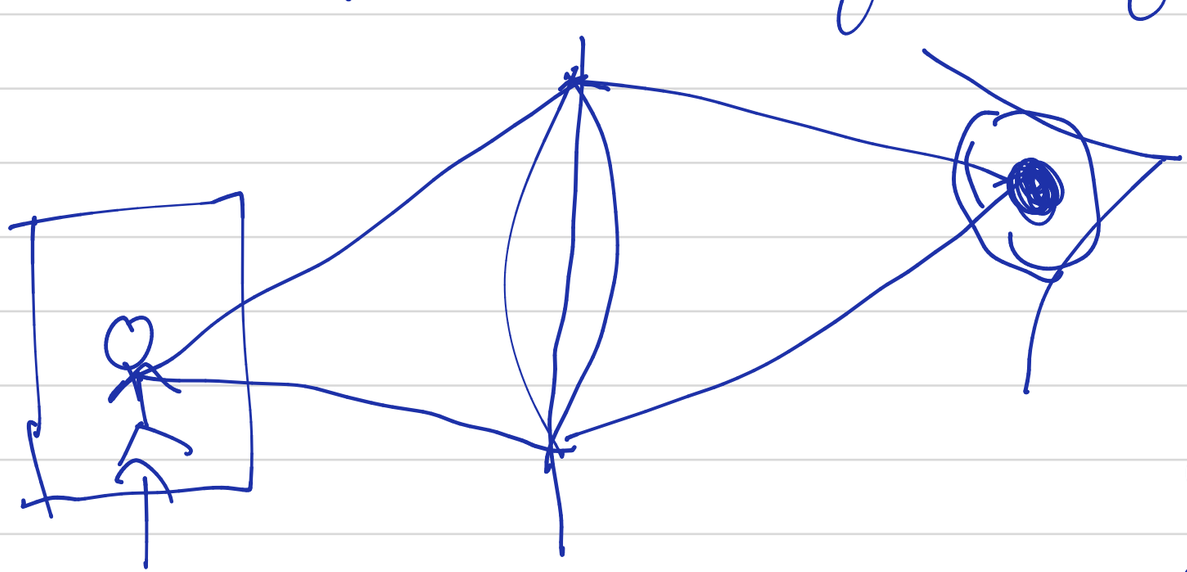


→ Cat Image / Dog Image

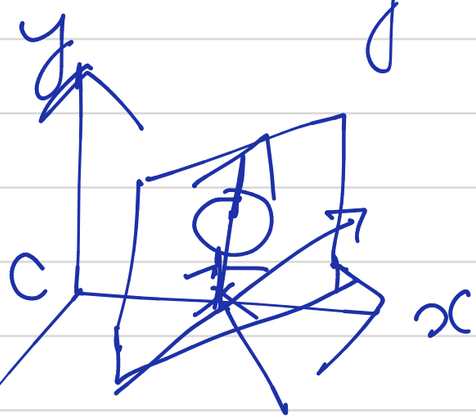
$N \times M \Rightarrow 3 \times 3$

Vector ⇒ a list of numbers in Space

→ For Human Eye it is easy but for Computer Vision Requires Mathematics, Light Intensity, Angle, Optics Maths



Human eye / Sensor / Lens.



Eg: for Computer Vision (figure)

Image As a Point in a Space.

$$A \Rightarrow \begin{bmatrix} 2 & 3 & 4 \\ 1 & 7 & 9 \end{bmatrix}$$

Then vector $A \Rightarrow$

$$\begin{bmatrix} 2 \\ 3 \\ 4 \\ 1 \\ 7 \\ 9 \end{bmatrix}$$

So, the Images can be Converted Similar to data points or vector.

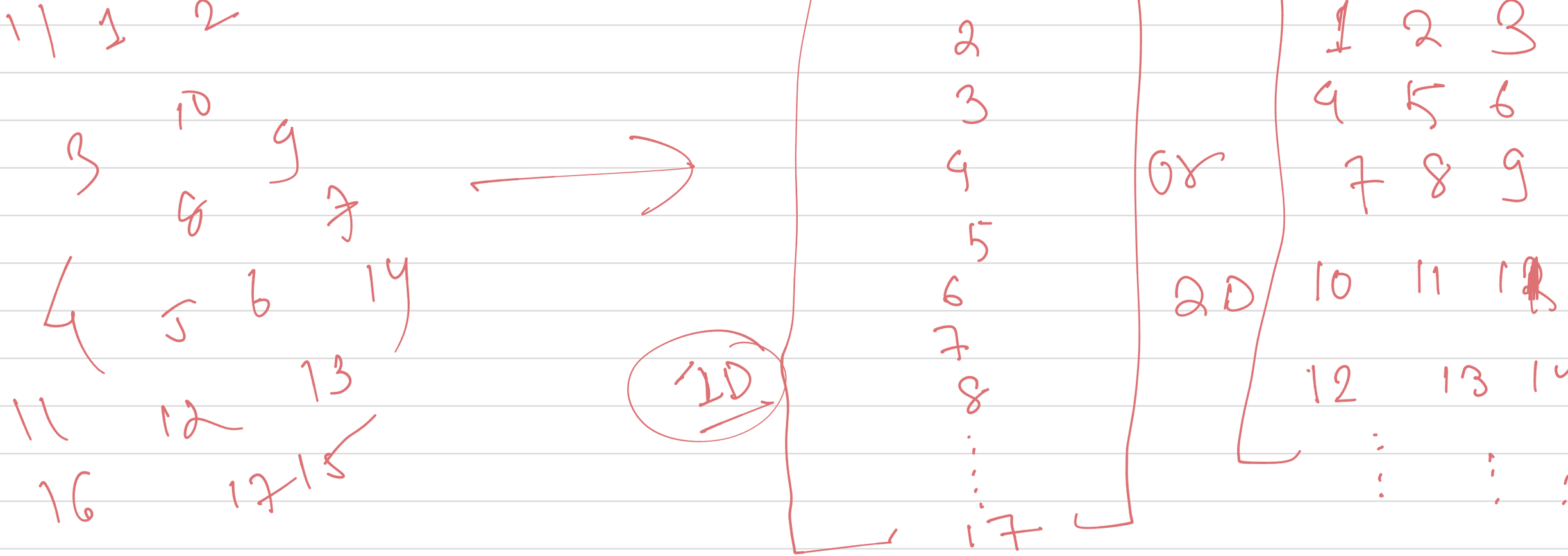
there could be 1D, 2D, 3D Dimensional vectors so,
for a vector (A), we can sketch following points

1) other kinds of data such as audio, language can be converted into an array of numbers

2) so, the data points is representation of the Input Data (images, video, Audio) can be features

3) Conceptually a "point" in a Space, but the structure of data and the Space varies for various problems

Think a Collection of numbers or the vectors and we compact them it becomes a vector encroached in a space.



Yes, we can now project the row vector = [. . . . .] or

n-Dimensional vector

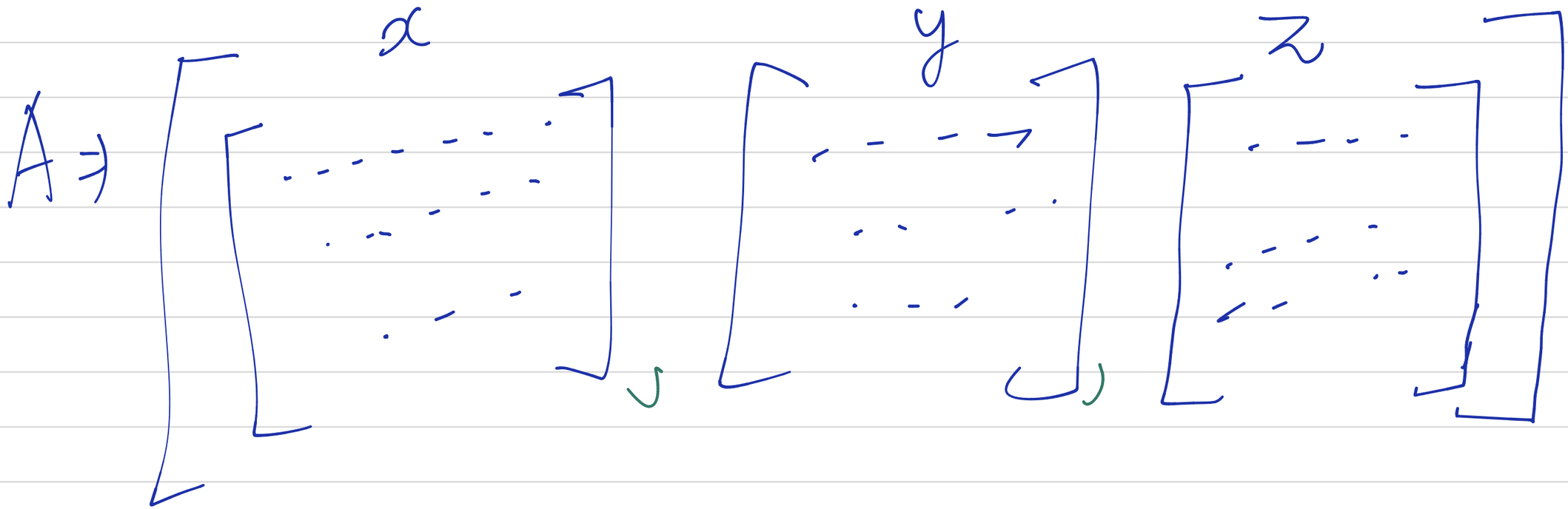
2D-vector of Suppose Image is $500 \times 500 \Rightarrow 2,50,000$ $(x_1 y)$ points

Vector (A) $\Rightarrow$ $\begin{bmatrix} x \\ y \\ . \end{bmatrix} \Rightarrow \begin{bmatrix} x_n \\ y_n \end{bmatrix} \Rightarrow \begin{bmatrix} 0,0 & \cdots & 0,499 \\ \vdots & & \\ 499,0 & \cdots & 499,499 \end{bmatrix}$

Here A[i][j] $\Rightarrow$ rows x column

if it is 3D then 500 x 500 x 500

it has 2D matrices depth or (the layer) + depth



A[i][j][k]

where i = depth 0 to d-1

j ⇒ row index 0 to r-1

k ⇒ column index 0 to c-1

Row Vector $v_n \Rightarrow [v_1, v_2 ---- v_n] \rightarrow (1 \times n)$ (vA)

→ 1 row, n column

Column Vector, $u_n \Rightarrow \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \rightarrow (n \times 1)$

→ n row, 1 column (Au)

So, finally we can multiply

$V \Rightarrow [1\ 2\ 3]$

$u \Rightarrow \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix}$

$V \cdot u \Rightarrow [1\ 2\ 3] \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix}$

$\Rightarrow 1 \cdot 5 + 2 \cdot 6 + 3 \cdot 7$

$\Rightarrow 38$

So, for the vector multiply $u \times v$

$u \Rightarrow \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$ and $v \Rightarrow \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$

So, $u \times v \Rightarrow \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$

$\Rightarrow \begin{bmatrix} 4.1 & 4.2 & 4.3 \\ 5.1 & 5.2 & 5.3 \\ 6.1 & 6.2 & 6.3 \end{bmatrix}$

$u \cdot v \Rightarrow \begin{bmatrix} 4 & 8 & 12 \\ 5 & 10 & 15 \\ 6 & 12 & 18 \end{bmatrix}$

→ Key Points.

→ Data (Input to the Model (Images, Videos))

→ Task (C classification, Regression, Clustering))

→ Model (CNN, RNN, Feed forward ANN)

→ Parameters (right Parameter findings)

→ What to optimize (loss function)

→ Model Parameters Optimiaqtion

→ Training process, Approarch

Goal: Build, Train and Optimize the ML Model that can work both Ideal and Real world-data, or the generaliaqtion of the Model that can give good accuracy and the prediction/output.

Software Development vs ML Products (ML Based Products)

→ In traditional software develop it is what it is specific features using if-else, switch case

→ but in facebook tagging friends, text translation, so it cannot be only if-else, face detection and tagline are highly non-static, dynamic and anonymous

→ function and Performance has blurrier boundary as compared to traditional Software development

ChatBot may or maynot work as expected but still in use but we don't have metric to define (when, where)

→ Generalization is the open and big Problems in model (ML Model)

→ Human in Loop.

→ Finding the right metric is not always obvious

→ ML Model will not completely replace or enable to automate itself till now so the Human aid (with the minimize effort) can or required to operate the "Model, a model can be learning algorithm or agent or mathematical model itself trained for a Specific intellectual tasks and behaviours

AI, Society & Ethics,

→ Using Facial Recognition for every person may not be suitable in every case,

Privacy issues

→ Diversity, Military Applications (Done Can Kill and Save both)

→ Bias, Surveillance

→ Winner takes all dynamics

→ Ethics and Morale on building the ML Models.

What do I learn ?

1) Linear Algebra (Computer Vision, Images/Videos)

Data → Self/Semi/Unsupervised

. Architectures → CNN's

Probability and Statistics →

Applications - Language NLP, Speech

Uses RNNs/LSTMs, Transformers

II Fundamental of ML

Applications → Biomedical Imaging

& Bioinformatics →

Dimensionality by Reduction

uses Convnets, Geometric Deep

Learning

#AI & Society → Networks →

Disentangling Latent Structure

& Reinforcement Learning

#AI & Ethics →

Operations, Research finance

& Insurance | Disaster

Risk Management