

Day 2 (4 Oct, 2024)

- Neural Networks are Superior than the traditional linear Regression or the linearly function
- Non-linearity function such as activation function (tanh, Sigmoid) are used in Neural Networks.

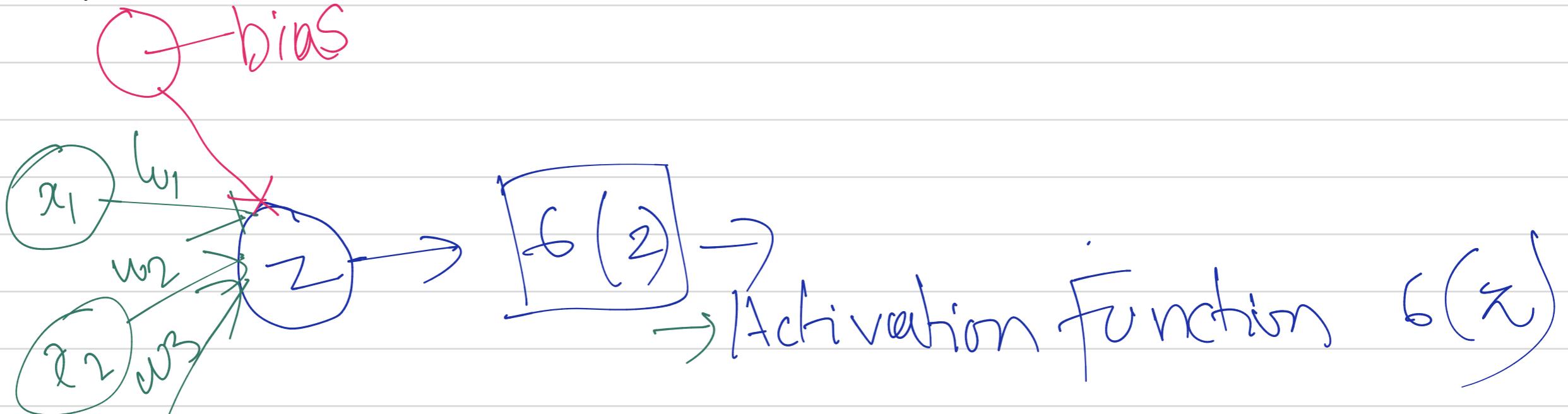
Sigmoid function with linear function

We have

$$Z = W_n x_n + \text{bias}$$

So if we sketch the figure with single neuron

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



A handwritten derivation of the sigmoid function. It starts with the equation $A = f(z) = \frac{1}{1 + e^{-z}}$. This is then equated to $A = \frac{1}{1 + e^{-(wx+b)}}$. A graph of the sigmoid function is shown, along with its derivative $\frac{d}{dx} \frac{1}{1 + e^{-x}} = \frac{e^{-x}}{(1 + e^{-x})^2}$.

→ 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 .

→ There are different methods to connect w weight of θ to the input and the neuron

and in 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 it into the parameter
 ω 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 $f(x_1, y_1, z)$
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

3

0 0 0 0 0 0
0 1 0 0 1 0
0 1 1 1 0 1
1 0 0 0 1 0

→ 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

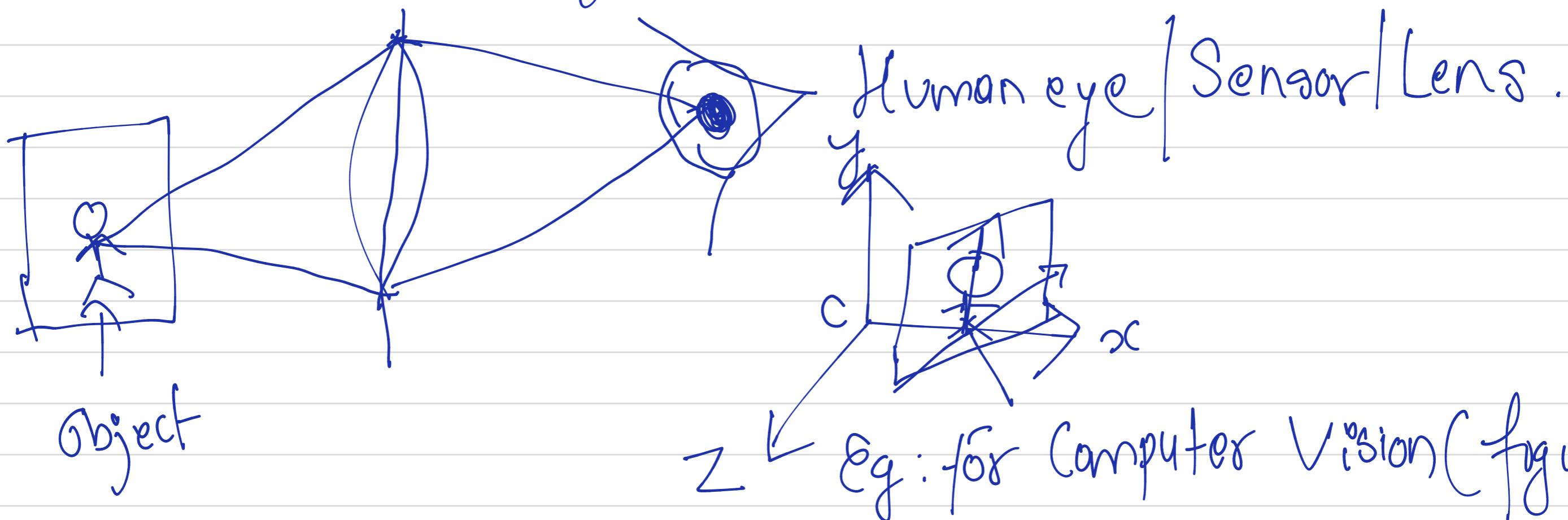


Image As a Point in a Space.

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

then vector $A \Rightarrow$

So, the images can be
converted similar to
data points or vectors.

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

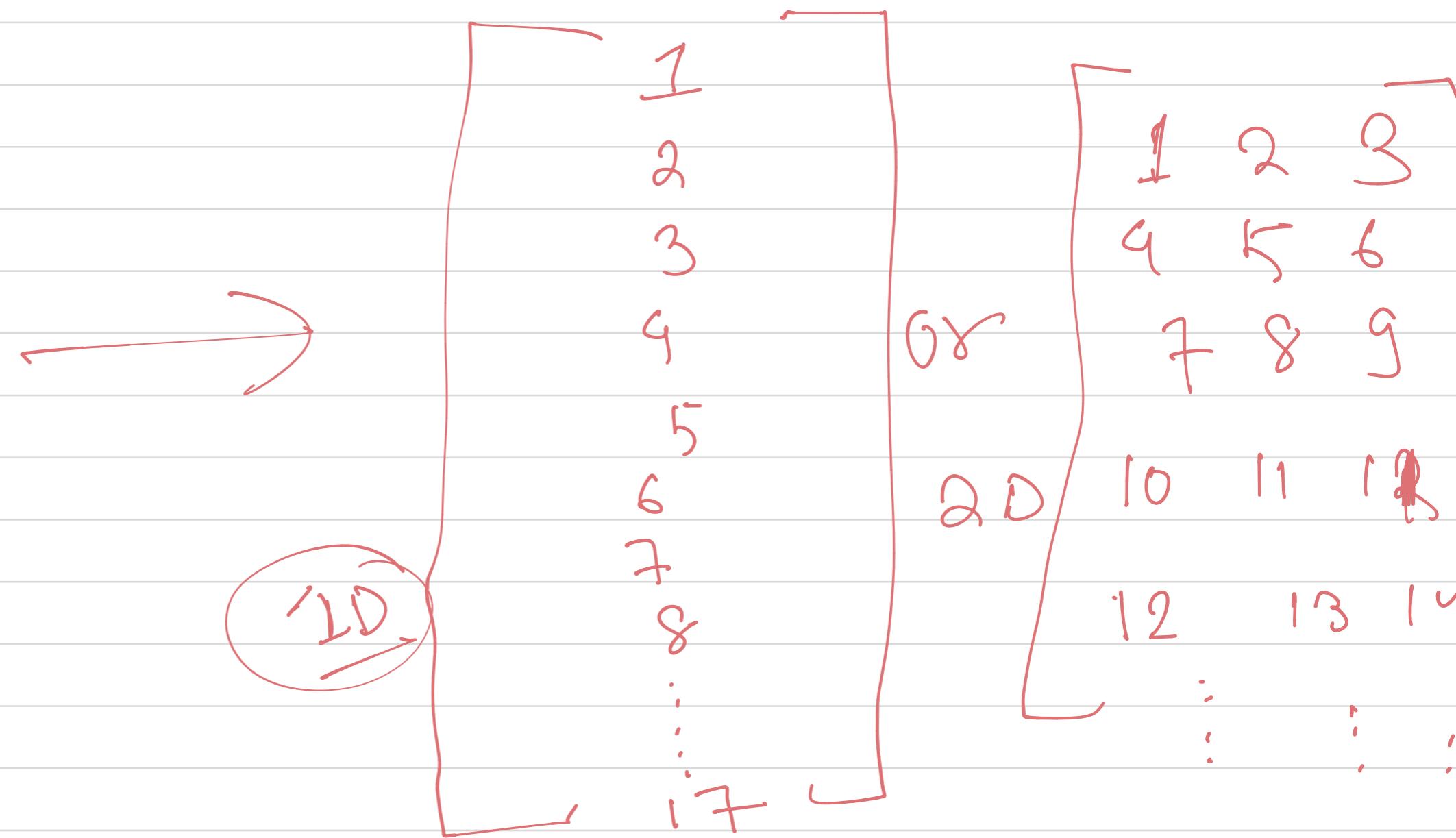
There Could be 1D, 2D, 3D Dimensional vector so,

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

Collection of
think about numbers or the vectors and we compact
them if becomes a vector enclosed in
a Space.

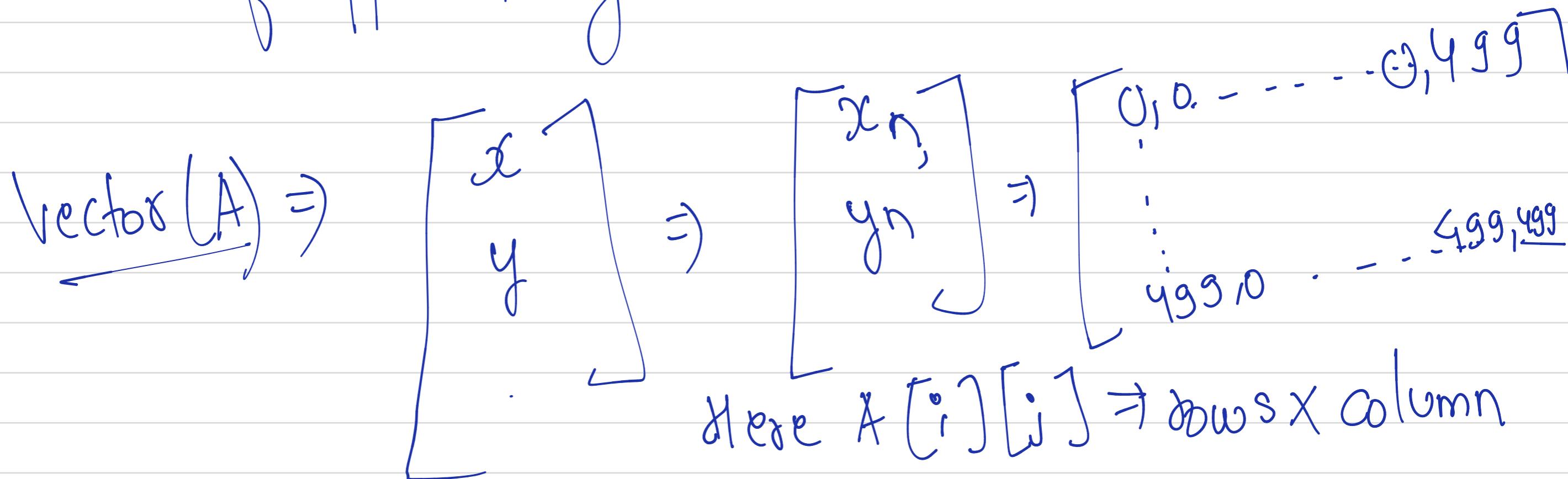
1 1 1 2
3 10 9
3 8 7
4 5 6 14
11 12 13
16 17 15



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

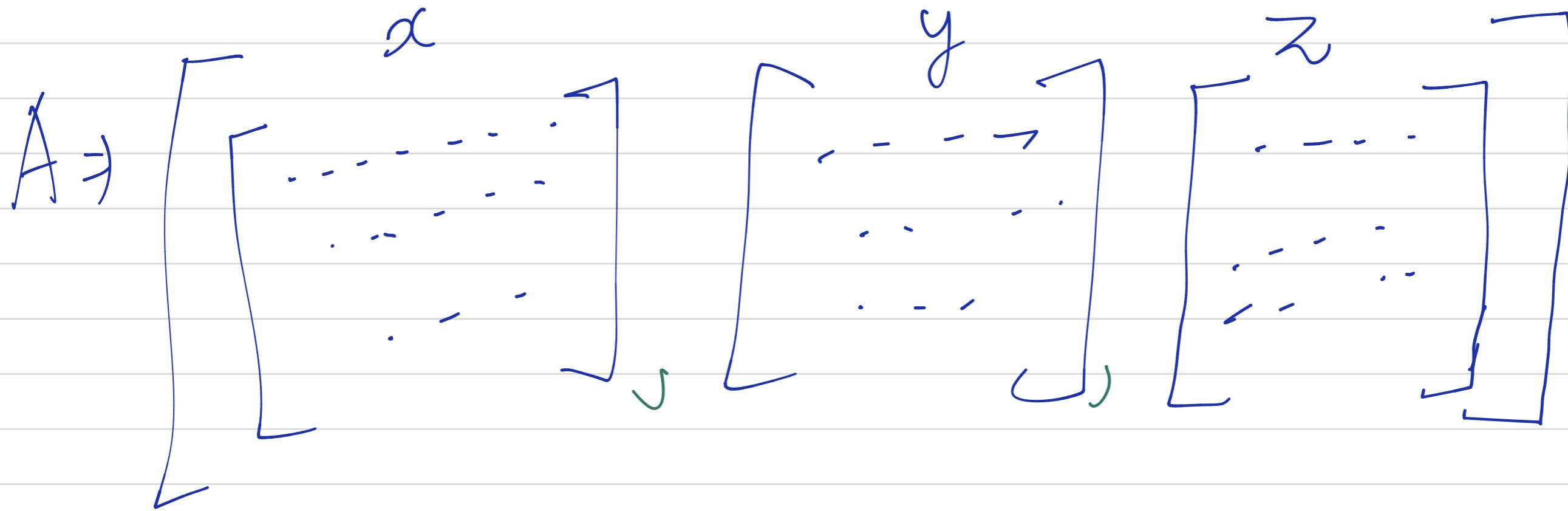
n-Dimensional Vector

2D-vector of Suppose Image is $500 \times 500 \rightarrow 250,000$ points
 (x_i, y_i)



if it is 3D then $500 \times 500 \times 500$

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



$A[i][j][k]$

where $i = \text{depth } 0 \text{ to } d-1$

$j = \text{row index } 0 \text{ to } j-1$
 $k = \text{Column index } 0 \text{ to } c-1$

Row Vector $v_n = [v_1, v_2, \dots, v_n] \rightarrow (1 \times n) (VA)$
 $\rightarrow 1 \text{ row, } n \text{ column}$

Column Vector, $u_n = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \rightarrow (n \times 1)$
 $\rightarrow n \text{ row, } 1 \text{ column } (AU)$

So finally we can multiply

$$v = [1 \ 2 \ 3]$$

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

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

$$\Rightarrow 1 \cdot 5 + 2 \cdot 6 + 3 \cdot 7 \\ \Rightarrow 38$$

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

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

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

$$\Rightarrow \begin{bmatrix} 4 \cdot 1 & 4 \cdot 2 & 4 \cdot 3 \\ 5 \cdot 1 & 5 \cdot 2 & 5 \cdot 3 \\ 6 \cdot 1 & 6 \cdot 2 & 6 \cdot 3 \end{bmatrix}$$

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

→ Key Points:

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

→ Task (Classification, Regression, Clustering)

→ Model (CNN, RNN, Feed forward ANN)

→ Parameters (right Parameter Findings)

→ What to optimize (Loss function)

+ Model Parameters Optimization
→ Training process, Approach

Goal: Build, Train and Optimize the ML Model that
Can work both Ideal and Real world-data,
or the generalization 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 no linear boundary
as compared to traditional Software development
- ChatBot may or may not 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 a role
to automate itself till now so the Human
aid (with the minimize effort) can
be 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 (Drones)

Can kill and save both

↪ Bias, Surveillance

→ Winner takes all dynamics
→ Ethics and Morale on building the
ML Models.

What do I learn?

Linear Algebra (Computer Vision)
(Images/Videos)

Data → Self/Semi/Unsupervised

• Architectures → CNN's

Probability and Statistics →

Applications - language NLP, Speech

uses RNNs | LSTMs, Transformers

Fundamentals of ML

Applications → Biomedical Imaging

bioinformatics →

Dimensionality by Reduction

USPS QNNs

Geometric Deep

Learning
~~Habit~~ & Society → Networks →

Disenchantment latent structure
Reinforcement learning

HAI & Ethics →

Operations, Research, Finance

Insurance, Disaster

Risk Management