#Day 11, oct-13-2024.

#Revision of Previous Days

1. Input to the Neural Network Model can be any object but object must be in data points that could be either matrix or vector.

Eg. $y = w_n x_n + b$ where b = bias

x = Input (data points)

and w is the weight of neuron

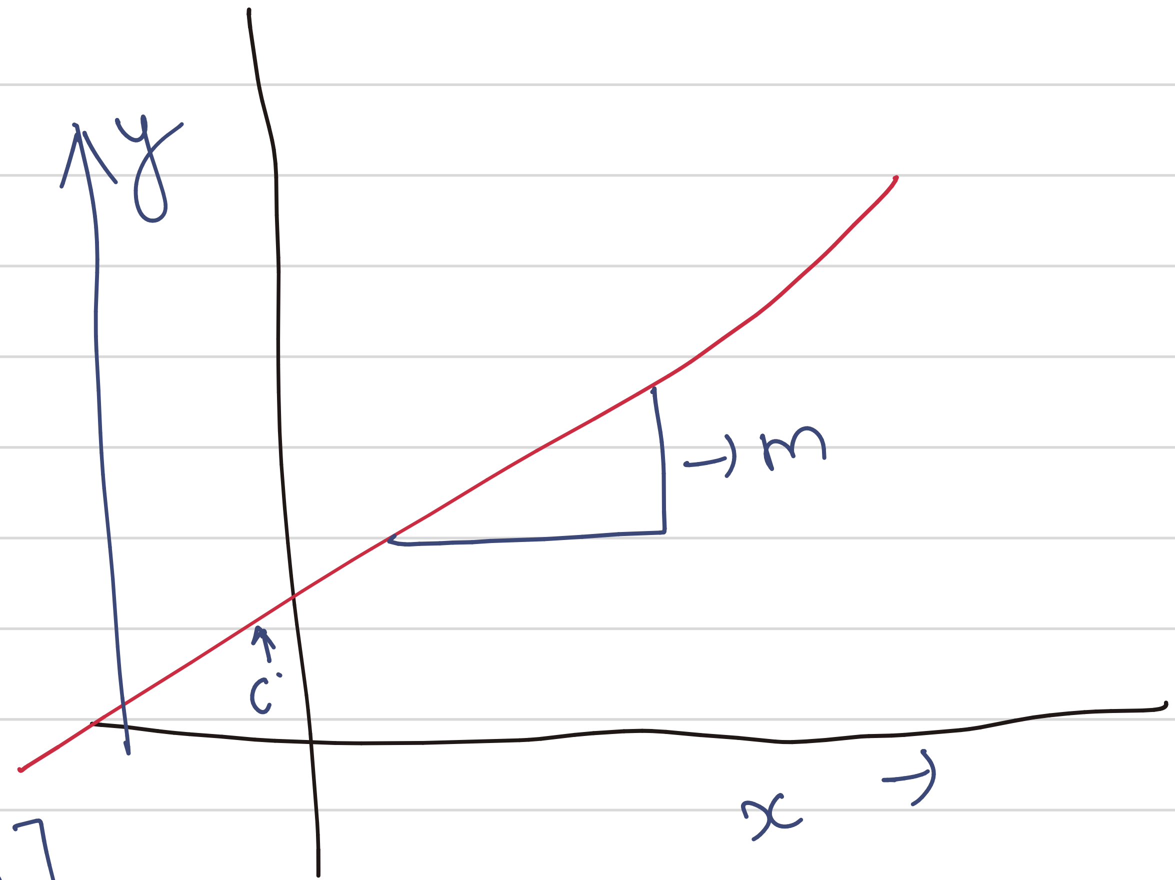
x could be $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

Stronger the connect optimized the weight. In the begining weight be any random.

So, the bias is the slope of a function or we can say gradient so,

$y = mx + c$ or

$(y = \beta_0 + \beta_1 x.)$



where x_n and w_n be

$w_n = [w_1, w_2, w_3 \dots \dots w_n]$

$x_n \Rightarrow \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and 'b' be the bias.

After applying Non-Linear function/Activation function $\sigma(z) \Rightarrow \frac{1}{1+e^{-z}}$

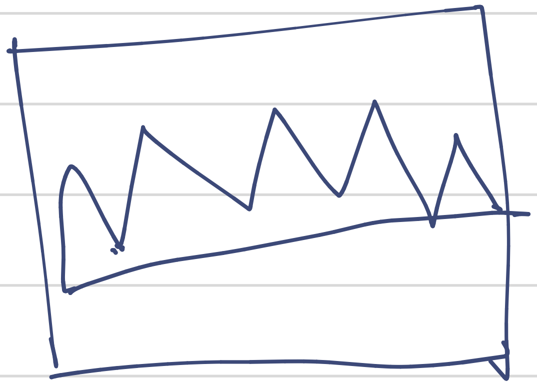
where $z = w_n x_n + b$.

(2) Images are 2D or 3D objects can be used as Object Datapoints in ANN (Neural Components)

$$\begin{matrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$$

Binary Coded Image Matrix

$\xrightarrow[\text{a image of}]{\text{is}}$



Suppose Original Image

[0 - 255 or 256 pixels value]

0 - 255 are also called Grayscale

So, 2D-Images are represented as $f(x,y)$ where x and y can be spatial co-ordinates and the intensity of $f(x,y)$ represents the pixels intensity or values.

Suppose $M \times N = 3 \times 3$ of 2D-Image $f(x,y)$ that can be

$$\begin{array}{cc} & 3 \\ 3 & \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \longrightarrow$$

Can be Cat/Dog (Although in real world Images can be 500x500 or more)

3D-Images Can be represented as $3 \times M \times N$ or Simply $f(x,y,z)$ and has 2D matrices depth or the 2-layer $500 \times 500 \times 500$

$$A \Rightarrow \left[\overset{X}{\begin{bmatrix} \cdots \end{bmatrix}}_{i} \times \overset{Y}{\begin{bmatrix} \cdots \end{bmatrix}}_{j} \times \overset{Z}{\begin{bmatrix} \cdots \end{bmatrix}}_{k} \right]$$

So, mathematically $A[i][j][k]$ where i = depth

$j \Rightarrow$ row index 0 to $r-1$

$K \Rightarrow$ col index 0 to $c-1$.

Let's understand the vector (row and Column vector).

(v) Vector in row form only = $[x_1, x_2, \ldots x_n]$

(u) Vector in only col form = $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

Both can be Scaled $(+, *)$

So, $\begin{bmatrix} 2 \\ 5 \end{bmatrix} * \begin{bmatrix} 2 & 5 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 \cdot 2 + 5 \cdot 5 \end{bmatrix}$

$v \times u$ $\Rightarrow 29$

Multiply ($\times$)

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

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

$$u \times v \Rightarrow \begin{bmatrix} 3 & 6 & 9 \\ 5 & 10 & 15 \\ 6 & 12 & 18 \end{bmatrix}$$

- → Dot product

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

$\Rightarrow 1 \cdot 5 + 2 \cdot 6 + 3 \cdot 7 \Rightarrow 5 + 12 + 21$

$\Rightarrow 38$

if. $u \times v \Rightarrow \begin{bmatrix} 5 \\ 6 \\ 8 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$

$$\Rightarrow \begin{bmatrix} 5\times1 & 5\times2 & 5\times3 \\ 6\times1 & 6\times2 & 6\times3 \\ 8\times1 & 8\times2 & 8\times3 \end{bmatrix} \Rightarrow \begin{bmatrix} 5 & 10 & 15 \\ 6 & 12 & 18 \\ 8 & 16 & 24 \end{bmatrix}$$

③ So, we must focus on the parameters, its optimization, and types of Model & tasks and also predicted error or the distance, training process, generalization and more.

④ Also AI with ethics, morale, cultural, Society, privacy and there will be always Human in loop (AI must be control)

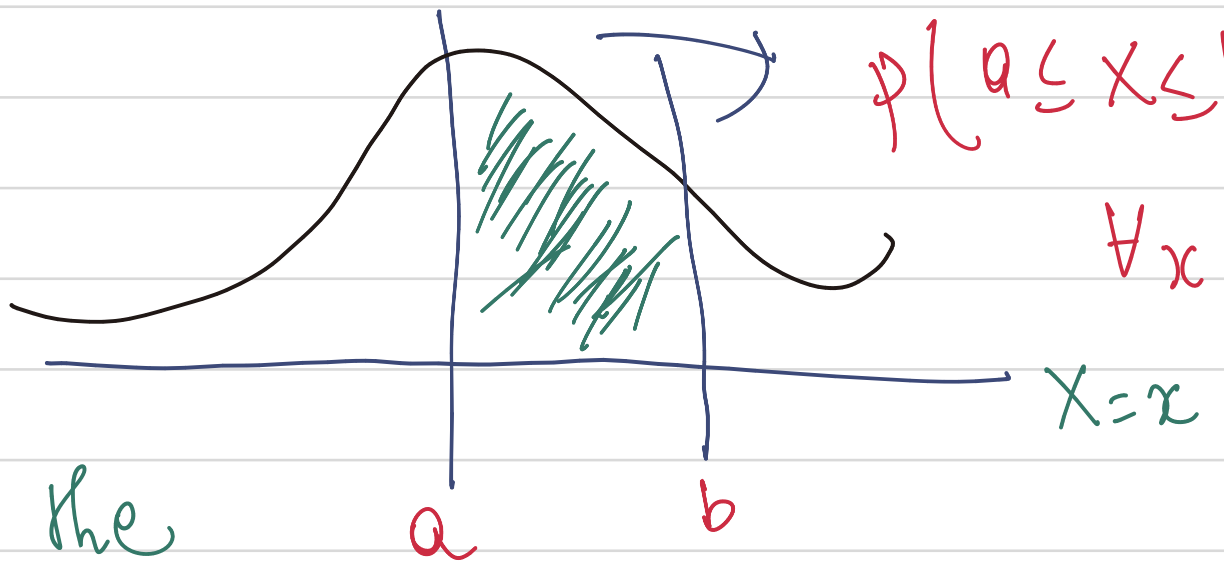
⑤ statistics is the study, analyzing, understanding, deriving pattern, drawing inferences from the empirical and data using statistical mathematic So, to find the middle value we need median for a given data or population or even sample. In ML, we depend and play with huge data so it is essential subject for AI and ML.

⑥ Probability is the likelihood of any event occurance. It handles the uncertainty of the real world. Continuous and Discrete are the main two types of data we take so probability for discrete is either 0 or 1 whereas for continuous is range [0,1] 0 to 1.

(7) Distributions is the probability of occuring all events descriptions. So, the different events follows different distributions. So, it is a graphical representations that represents all events of probability. Gaussian, Normal, Exponential, Poisson Distributions are some examples.

(8) Probability can be only betwee 0 to 1 or either 0 or 1 so to represent discrete outcome for discrete input probability we have PMF (Probability Mass function) $\forall x \in X$ $0 \leq p(x) \leq 1$ then the integral exists $\sum_{x} p(x) = 1$

9 PDF (Probability Density function) takes X continuous random variables and x is a random variables, Here probability lies in between $p(a \le X \le b)$ So, it is the most useful and resemblance to the real world.



$$p(a \le X \le b) = \int_a^b f(x)\,dx$$

$$\forall x \in X : p(x) \ge 0$$

The figure itself is the normal distribution (describes the probability of all events ocurrance in Space (Sample Space))

Where state Space can be Head or tail.

10 Bernoulli Distribution → Instead of showing $p(x)$ as 0 or 1 or [0,1] we can use PMF (need not explicitly mention Head or tail)

So just putting x we can have outcome either 0 or 1.

$$p(X=x) = \phi^x (1-\phi)^{1-x} \quad \text{PMF}$$

$$E_X[x] = \phi \rightarrow \text{Expectation}$$

$$Var_X[x] = \phi(1-\phi) \rightarrow \text{variance}$$

for $x=0$

$$p(X=0) = \phi^0 (1-\phi)^{1-0}$$

$$\Rightarrow 1-\phi \qquad \text{where } \phi \text{ is a parameter}$$

(ii) Gaussian Distribution follows symmetry, mean (μ) '0' and peakedness and is used to represent distribution for continuous probability

$$p(x) \Rightarrow \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma}}$$

where x = random variable

$\sigma \Rightarrow$ std. deviation

So, the Gaussian distribution looks all

(12) Joint Probability Distribution describes the probability of multiple events occuring simultaneously, can be continuous and discrete both. $P(X=x_i, Y=y_j)$ JPD for discrete & $P(a \le X \le b, c \le Y \le d) =$ for continuous. As $P(x,y) \Rightarrow \int_c^d \int_a^b f(x,y) \cdot dx \cdot dy$

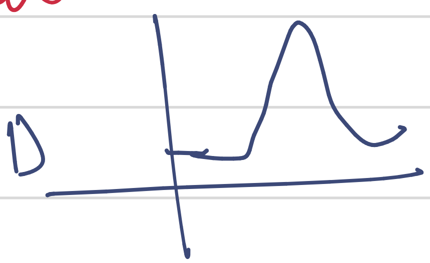
$$P(a \le X \le b) = \int_a^b f(x) \cdot dx$$

JPD — Coin toss, Dice Roll — Simultaneously

(13) Marginal probability gurontee one outcome (p') among two outcome

for discrete $P(X=x) = \sum_{y} P(X=x, Y=y)$

for Continous $P(X=x) = \int f(x,y) \cdot dy$ and we can use Heat map for this. It is closely related to Conditional probability $P(X|Y=1) = \frac{P(X \cap Y)}{P(Y)}$ and it is slicing of 1x4D



(14) Multi-variate Gaussian distribution generalizes the normal distribution, more flexible representation of complex, multi-dimensional data, model relationship between multiple random variables in a Continuous Space.

(15) Suppose we have multiple probability events then we need Chain Rule that breaks down Complex probabilities, multi-event systems and nested functions.

$$p(A, B, C) = p(A) \cdot p(B/A) \cdot p(C \mid A, B)$$

(16) Expectation or Expected Mean is the average outcome of random variable. Represents the weighted average of all possible values that a random variable can take.

$$E_x[g(x)] = \sum_{x \in X} g(x) \cdot p(x)$$

$$E_x[x] = \int_{-\infty}^{\infty} x\, p(x) \cdot dx$$

Eg: 1, 2, 3

$$E(x) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3}$$

$$\Rightarrow \frac{1+2+3}{3} \Rightarrow 2$$

Expectation of Uniform Distribution

$$f(x) = \frac{1}{3} \text{ for } 2 \leq x \leq 5$$

$$E(x) = \int_2^5 x \cdot \frac{1}{3}\, x \Rightarrow \int_a^b x\, f(x) \cdot dx$$

$$\Rightarrow \frac{1}{3} \int_2^5 x \cdot dx$$

$$\int x \cdot dx \Rightarrow \frac{x^2}{2}$$

So, $\frac{1}{3}\left[\frac{x^2}{2}\right]_2^5$

$$\Rightarrow \frac{1}{3}\left[\frac{5^2}{2} - \frac{2^2}{2}\right]$$

$\Rightarrow \frac{21}{6} \Rightarrow 3.5$ which is $E(X) = 3.5$ expected value of X for a Continuous random variable - mean.

17 **Independence:** when two events occurs (X) and (Y) independently

$$P(X=x, Y=y) = P(X=x) \cdot P(Y=y)$$

18 **Variance:** Spread of a data-sets or value around the μ and show how closely relative to mean.

$$Var(x) = E(x^2) - [E(x)]^2$$

19 **Co-variance:** How two random variables are related and dependent upon eachother (univariate random variable) $X, Y \in R$

$$\mathrm{Cov}[x,y] = E[xy] - E[x] \cdot E[Y]$$

20) Co-variance Matrix: Square Matrix of set of n-random variables where diagonal elements represents variance ($Var(X_i)$) and

$$\mathrm{Cov}(X,Y) = E[(X - E(X))(Y - E(Y))]$$

Matrix is

$$\Sigma = \begin{pmatrix} Var(X_1) & \cdots & Cov(X_1, X_2) & \cdots & Cov(X_1 \cdots X_n) \\ Cov(X_2, X_1) & \cdots & Var(X_2, X_2) & \cdots & Cov(X_2, X_n) \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ Cov(X_n, X_1) & \cdots & Cov(X_n, X_2) & \cdots & Var(X_n) \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} var(X) & Cov(X,Y) \\ Cov(X,Y) & var(X) \end{pmatrix}$$

Square 2×2 matrix with n-random Variables x_1, x_2, x_n

(20) Linear Algebra study of Matrices, determinants, vectors, vector Space, Linear transformation.

(21) Vector → a collection of number $x_1, x_2 \cdots x_n$ row vector = $[\cdot - - - \cdot]$

Column vector $\begin{bmatrix} \vdots \\ \vdots \end{bmatrix}$, denoted by $\vec{v}$.

(22) if two or vector can be multiplied or added by Scalars (real or Complex) with rules then it is vector Space.

(23) Linear transformation : A function maps vector from one Space to another while preserving (Scalar and Addition).

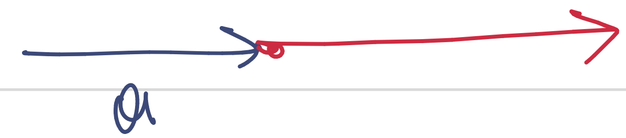
(24) Matrices are the collection of set of ordered vectors. It can represents system of equations, pixels of images, represents graph, linear transformations, dimensionality Reduction.

(25) Vector Scaling means multiplication or Sketching, Squishing vector only in magnitude. So, the Span of vectors means the set of all linear vectors.

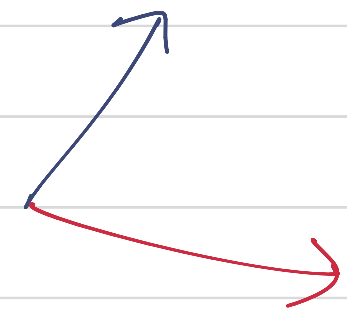
i.e $a\vec{v} + b\vec{u}$

(26) dependent upon each other vectors are Linearly Dependent

$\vec{u} = a\vec{v} + b\vec{w}$ So,



27 Linearly independent are set of vectors that are not dependents Upon each other



Both Linearly Independent & dependent can span but the vectors may not span anymore in Some Special cases in 2D or 3D Space.

28 Basis of a Vector Space is a Collection of Linearly Independent vectors that span the full vector Space.

29 Vector Products are Hadamard products (element-by-element),

Outer product, Inner Product.

(30) Hadamord Product (CNN) $\begin{bmatrix} a_1 \\ b_1 \end{bmatrix} \odot \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} \Rightarrow \begin{bmatrix} a_1 a_2 \\ b_1 b_2 \end{bmatrix}$

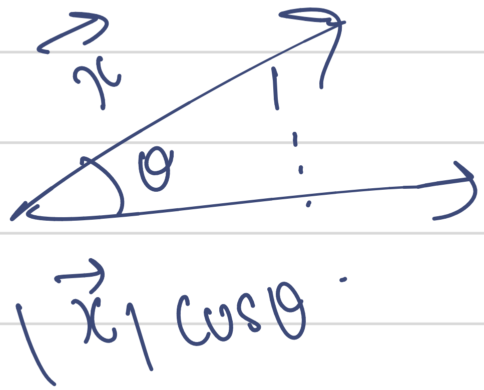
(31) Outer product $u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$, $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$

$1 \times N$ $\qquad$ $N \times 1$

$$u \times v = A = \begin{bmatrix} u_1 v_1 & u_1 v_2 & \cdots & u_1 v_n \\ u_2 v_1 & u_2 v_3 & \cdots & v_2 v_n \\ \vdots & \vdots & & \vdots \\ u_n v_1 & u_m v_2 & \cdots & u_m v_n \end{bmatrix}$$

(32) Dot product are applied ANN. $\vec{x} \cdot \vec{y} \Rightarrow x_1 \cdot y_1 + x_2 \cdot y_2 + x_n y_n$

(33) Geometric Dot Product Interpretation.



$$\vec{x} \cdot \vec{y} = |\vec{x}|\,|\vec{y}| \cdot \cos(\theta)$$

(34) Eigenvectors & Eigen values are those that fulfill $Ax = \lambda x$ where λ is eigenvalue and x is eigen vector that changes the magnitude but not the direction. It can help in dimension-reduction, represent data as a matrix using eigen decomposition (matrix decompose into eigenvectors and finding their scale vectors) and for PDA (principle Direction of Data)