

MATHEMATICS For MACHINE LEARNING

⇒ LINEAR ALGEBRA

Notations →

A, B, C

$\vec{u}, \vec{v}, \vec{w}$

A of size $m \times n$ or $(m \times n)$

A^T

$\vec{v}^T$

A^{-1}

$\det(A)$

AB

$\vec{u} \cdot \vec{v} ; \langle \vec{u}, \vec{v} \rangle$

$\mathbb{R}$

$\mathbb{R}^2$

$\mathbb{R}^n$

$\vec{v} \in \mathbb{R}^2$

$|\vec{v}|_1$

$|\vec{v}|_2 ; |\vec{v}| ; \|\vec{v}\|$

→ Capital letters represent matrices

→ lowercase letters represent vectors

→ matrix A has m rows & n columns

→ The transpose of matrix A

→ The transpose of vector $\vec{v}$

→ The inverse of matrix A

→ The Determinant of matrix A

→ matrix multiplication of matrices A & B .

→ dot product of vectors $\vec{u}$ & $\vec{v}$.

→ The set of real numbers, e.g. $0, -0.67, 2$

→ The set of two-dimensional vectors, e.g. $\vec{v} = [1 \ 3]^T$

→ The set of n -dimensional vectors.

→ Vector $\vec{v}$ is an element of $\mathbb{R}^2$

→ L1-norm of Vector

→ L2-norm of Vector

$T: \mathbb{R}^2 \rightarrow \mathbb{R}^3 ; T(\vec{v}) = \vec{w} \rightarrow$ transformation T of a vector $\vec{v} \in \mathbb{R}^2$ into the vector $\vec{w} \in \mathbb{R}^3$

Application of linear algebra in machine learning

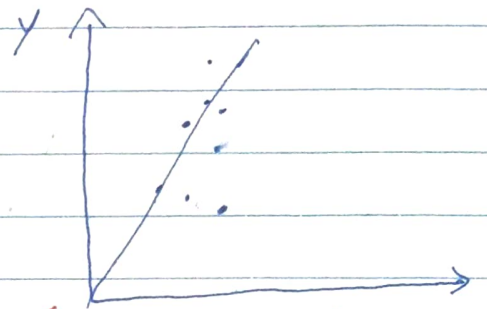
→ Linear regression :-

it is Supervised machine learning algo

- It finds the line of best fit for the data on the graph.
- The relationship between input & output is linear.

$$m \times (\text{Input}) + b = (\text{Output})$$

$$m x + b = y$$



→ Not a correct diagram
just for intuition.

↳ In machine learning

y = output

m = weight represented by (w)

x = Input

b = bias

⇒ But what if there are two inputs or more,

→ Then we add another variable & a weight onto that variable.

$$w_1 x_1 + w_2 x_2 + b = y$$

→ If you graph this equation now it will not be in 2d like given above it will be in 3d, the (higher the number of input variables higher the number's of Dimension) & it will form a plane not line.

⇒ what if you have a lot of variables then we do

$$\rightarrow w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_nx_n + b = y$$

→ In this algorithm you already know the value of (x) which is input & (y) which is output, you just need to find the ~~the~~ value of (w) which is weight & (b) which is bias, the correct value of (w) & (b) will satisfy the equation or almost satisfy the equation.

⇒ Your data will have multiple values or can say equations, so for multiple equations we can write:-

$$w_1x_1^{(1)} + w_2x_2^{(1)} + w_3x_3^{(1)} + \dots + w_nx_n^{(1)} + b = y^{(1)}$$

$$w_1x_1^{(2)} + w_2x_2^{(2)} + w_3x_3^{(2)} + \dots + w_nx_n^{(2)} + b = y^{(2)}$$

$$w_1x_1^{(3)} + w_2x_2^{(3)} + w_3x_3^{(3)} + \dots + w_nx_n^{(3)} + b = y^{(3)}$$

$$\vdots$$
$$w_1x_1^{(m)} + w_2x_2^{(m)} + w_3x_3^{(m)} + \dots + w_nx_n^{(m)} + b = y^{(m)}$$

equations

⇒ Notice that $(1), (2), (3) \& (m)$ here are denoting different equations & are called superscript's, these are not exponent's.

⇒ Our goal here is to solve every equation or to get as close to solving as possible with ~~same~~ weight's & biases being the same.

⇒ The equations on the previous page can also be represented by the matrix method like this.

$$[w_1, w_2, w_3 \dots w_n] \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \dots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \dots & x_n^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & x_3^{(m)} & \dots & x_n^{(m)} \end{bmatrix} + b =$$

$$[y^{(1)}, y^{(2)} \dots y^{(m)}]$$

$$W \cdot X + b = \hat{y}$$

System of Sentences

System 1

The dog is black
The cat is orange
(Complete) (non-singular)

System 2

The dog is black
The dog is black
(Redundant) (singular)

System 3

The dog is Black
The dog is white
(Contradictory) (singular)

Here, System 1 has 2 sentences & they both give a unique information, That's why it is called complete System

System 2 has 2 sentences but both the sentences carries same information so it is called redundant System

System 3 has 2 sentences but both the sentences contradict each other so it is called Contradictory System.

Here, the Complete System is called "Non-singular"
& others are called singular system

What is a linear equation?

⇒ an equation where the variables are only added or subtracted
like

linear $\Rightarrow 4a + 5b = 20$

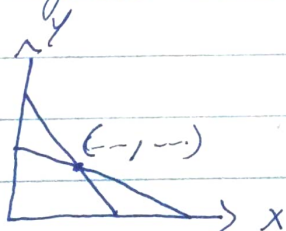
Non-linear $\Rightarrow a^2 + b^2 = 20$

$\sin(a) + b^2 = 35$

$2^a - 3^b = 54$

System of equations as lines & planes

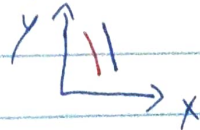
for system 1, lines cross each other at a specific point & that gives us the correct answers of x & y .



for system 2, lines ~~cross~~ ^{are parallel to} each other at a specific point & that gives us the correct ans & they doesn't cross

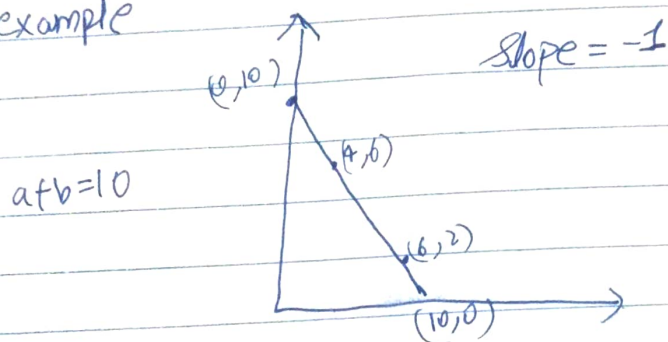
for system 2, lines overlap each other & hence they are crossing each other at infinitely many points so they will have infinite solutions.

for system 3, The lines simply don't cross each other & hence there is no solⁿ



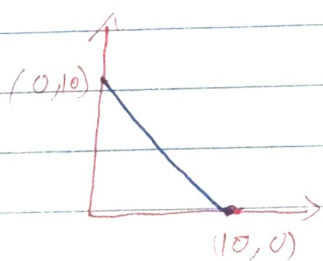
Slope $\therefore$ it is the ratio of rise over run,

example



Here slope is minus one because for every unit you move to right, the line moves one unit down.

Y Intercept $\rightarrow$ It is the height of the intersection between the line & Y axis.



Here, the Y intercept is 10, meaning the point of contact between line & Y-axis is 10.

$\Rightarrow$ A linear eqn with two variables corresponds to a line in ~~space~~ the plane,

$\Rightarrow$ A linear eqn with 3 variable will correspond to a plane in space.

⇒ The constant doesn't matter while determining the system is singular or not

$$\begin{matrix} a+b=0 \\ 2a+2b=0 \end{matrix} \rightarrow \text{constant's}$$

we can put constant's as zero to tell if a system of eq's is singular or not.

Singular vs non-singular matrix

System of equations as matrices

$$\begin{matrix} \text{System 1} & a & b \\ a+b=0 & 1 & 1 \\ a+2b=0 & 1 & 2 \end{matrix}$$

$$\begin{matrix} \text{System 2} & a & b \\ a+b=0 & 1 & 1 \\ 2a+2b=0 & 2 & 2 \end{matrix}$$

if a system of equation has is singular then its corresponding matrix is also singular & if a system of eqⁿ is non-singular then its corresponding matrix is also non singular

Linear dependence & Independence

⇒ The Rows which are linearly dependent don't have unique solⁿs & Rows which are linearly Independent have unique solutions.

$$\begin{matrix} \text{Non-singular} & a & b \\ a+b=0 & 1 & 1 \\ a+2b=0 & 1 & 2 \end{matrix}$$

↳ Second row isn't the multiple of first row.

$$\begin{matrix} \text{Singular} & a & b \\ a+b=0 & 1 & 1 \\ 2a+2b=0 & 2 & 2 \end{matrix}$$

↳ Here second row is multiple of first row
and two

If the sum or average sum of linear rows is equal to the third one it is linearly dependent.

The determinant of a matrix

If the determinant is zero the matrix is singular
& If it is non-zero then it's non-singular

Solving a redundant equation.

$$\left. \begin{array}{l} a+b=10 \\ 2a+2b=20 \end{array} \right\} \begin{array}{l} \text{Here} \\ \text{let, } a=x \\ b=10-x \end{array}$$

~~ty~~

Matrix Row Reduction or gaussian elimination

manipulating the rows of matrix to get the solution.

Original system

$$5a + b = 17$$

$$4a - 3b = 6$$

Intermediate system

$$a + 0.2b = 3.4$$

$$b = 2$$

Solved system

$$1a + 0b = 3$$

$$0a + 1b = 2$$

Original matrix

$$\begin{bmatrix} 5 & 1 \\ 4 & -3 \end{bmatrix}$$

Intermediate matrix

$$\begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix}$$

Row echelon
form

Diagonal matrix

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

Reduced row
echelon form

when the d_i elements
underneath diagonal are 0.

Row echelon form :-

1	2	3	5
0	1	4	6
0	0	0	0
0	0	0	0

⇒ In Row echelon form, below the diagonal everything must be zero, & above the diagonal anything getting succeeded by 1 can be any number, but anything getting succeeded by 0 has to be zero.

Row operations that preserves singularity.

The rank of a matrix

Compressing an Image works on the principle of "Reducing rank".

⇒ Pixelated images are matrices & the rank of a matrix is related to the amount of space that is needed to store that corresponding image.

"SVD (Singular value Decomposition) technique is used to reduce the rank of a matrix while changing it as little as possible"

The amount of information a system carries defines the "rank" of the system

System 1

The dog is black
The cat is orange
Rank = 2

System 2

The dog is black
The dog is black
Rank = 1

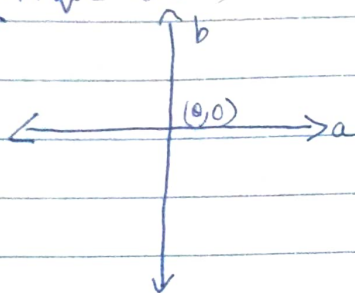
System

The dog is - -
The dog is - -
Rank = 0

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

Rank = 2

Dimension of
Solution Space = 0
("unique solⁿ")

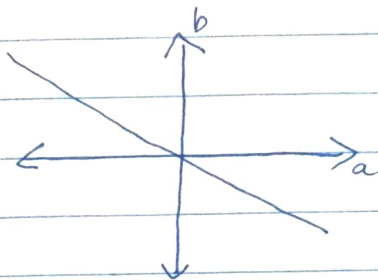


"Here you get a
unique solⁿ"

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$$

Rank = 1

Dimension of
Solⁿ Space = 1

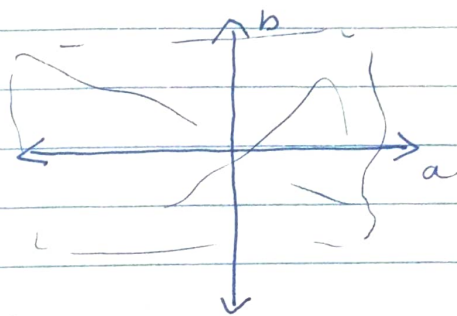


"Here you get
a line of solⁿ"

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Rank = 0

Dimension of Solⁿ
Space = 2



"Here any point
in the plane can
be a solⁿ"

"A matrix is non-singular if and only if it
has full rank" (in this case it is 2)
specific

Row echelon form

$$\Rightarrow \begin{bmatrix} 5 & 1 \\ 4 & -3 \end{bmatrix}$$

original matrix

Step 1 $\Rightarrow$ divide each row by the leftmost coefficient

$$\Rightarrow \begin{bmatrix} 1 & 0.25 \\ 1 & -0.75 \end{bmatrix}$$

Now $\Rightarrow$ Subtract the rows & use it as the bottom most row for the new matrix.
& write the first row as it is.

$$\Rightarrow \begin{bmatrix} 1 & 0.25 \\ 0 & -0.5 \end{bmatrix}$$

$\Rightarrow$ Divide the second row by the leftmost non-zero coefficient.

$$\begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix}$$

The number of ones in the diagonal will be the no. of rank

Rank in bigger matrix

$$\begin{bmatrix} 2 & * & * & * & * \\ 0 & 1 & * & * & * \\ 0 & 0 & 3 & * & * \\ 0 & 0 & 0 & -5 & * \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

⇒ The number of pivots is equal to the number of rank

⇒ Pivot is the leftmost non-zero no of a row

⇒ from top to bottom pivot

should go from left to right

meaning the pivot of the second row has to be at the right side of the pivot of first row.

Reduced Row echelon form

⇒

Row echelon form

$$\begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix}$$

→

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

⇒ leave the bottom row as it is

⇒ Take the bottom row multiply it by "0.2"
(because we want to get rid of 0.2)

⇒ & Subtract it from the first row & you are done.

for bigger matrix to be in an ^{reduced} row echelon form.

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

Rank 5

- * It has to be in row echelon form
- * Each pivot is a 1
- * any no. above a pivot is 0.
- * Rank of the matrix is no of pivot's.

How to ^{go to} reduced row-echelon form, for a bigger matrix

ex $\Rightarrow$ Row echelon form

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$$

Step 1 $\Rightarrow$ get ~~any~~ anything above pivot's to 0.

So $\Rightarrow$ Subtracting 2 times the second row from the first one.

$$\begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$$

Now, $\Rightarrow$ Subtracting 4 times of row three from row 2 & -5 times of row 3 from row one.

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

$\rightarrow$ reduced row echelon form

The gaussian Elimination Algorithm

Solving a system of eqⁿ.

$$\begin{aligned}\Rightarrow 2a - b + c &= 1 \\ \Rightarrow 2a + 2b + 4c &= -2 \\ \Rightarrow 4a + b &= -1\end{aligned}$$

$$\rightarrow \left[\begin{array}{ccc|c} 2 & -1 & 1 & 1 \\ 2 & 2 & 4 & -2 \\ 4 & 1 & 0 & -1 \end{array} \right]$$

$\Downarrow$

$$\left[\begin{array}{ccc|c} 2 & -1 & 1 & 1 \\ 2 & 2 & 4 & -2 \\ 4 & 1 & 0 & -1 \end{array} \right]$$

Augmented matrix

$\Rightarrow$ In Augmented matrix you add another column where you put all of your constant values

step $\Rightarrow$ Select top left cell as pivot & change it to zero

The operations will also be applied on the constant columns too

~~step~~ Dividing the first row by 2 to get 1.

$$\left[\begin{array}{ccc|c} 1 & -1/2 & 1/2 & 1/2 \\ 2 & 2 & 4 & -2 \\ 4 & 1 & 0 & -1 \end{array} \right]$$

step $\Rightarrow$ Now change all the values below the pivot to zero.

Subtracting 2 Row 1 from R_2 to get R_2
 $R_2 \leftarrow R_2 - 2R_1$

$$\left[\begin{array}{ccc|c} 1 & -1/2 & 1/2 & 1/2 \\ 0 & 3\frac{1}{2} & 3\frac{1}{2} & -3 \\ 4 & 1 & 0 & -1 \end{array} \right]$$

$$R_3 \leftarrow R_3 - 4R_1$$

$$\left[\begin{array}{ccc|c} 1 & -1/2 & 1/2 & 1/2 \\ 0 & 3 & 3 & -3 \\ 0 & 3 & -2 & -3 \end{array} \right]$$

$\Rightarrow$ Now doing the exact same thing for row 2 & 3 to get

$$\left[\begin{array}{ccc|c} 1 & -1/2 & 1/2 & 1/2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 0 \end{array} \right]$$

Now, using Back Substitution convert all the values above pivots to zero.

$$R_2 \leftarrow R_2 - R_3$$

$$\& R_1 \leftarrow R_1 - 1/2 R_3$$

$$\left[\begin{array}{ccc|c} 1 & -1/2 & 0 & 1/2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{array} \right]$$

* The Back Substitution for third row is done now let's do it for second row.

$$R_1 \leftarrow R_1 + 1/2 R_2$$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{array} \right]$$

⇒

reduced row echelon form

So,
$$\begin{aligned} 1a + 0b + 0c &= 0 \rightarrow a = 0 \\ 0a + 1b + 0c &= -1 \rightarrow b = -1 \\ 0a + 0b + 1c &= 0 \rightarrow c = 0 \end{aligned}$$

⇒ The square part of the augmented matrix has 1 in its diagonal & it's an identity matrix

Checking if it has infinitely many or no solutions.

look at the column of constants

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 5 \\ 2 & 4 & 5 & 1 \\ 3 & 5 & 4 & 6 \end{array} \right]$$

after row reduction →

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 5 \\ 0 & 0 & -7 & 9 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

$$0a + 0b + 0c = 0$$

↳ Infinitely many sol's

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 5 \\ 2 & 4 & 5 & 1 \\ 3 & 5 & 4 & 10 \end{array} \right]$$

after row reduction →

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 5 \\ 0 & 0 & -7 & 9 \\ 0 & 0 & 0 & 4 \end{array} \right]$$

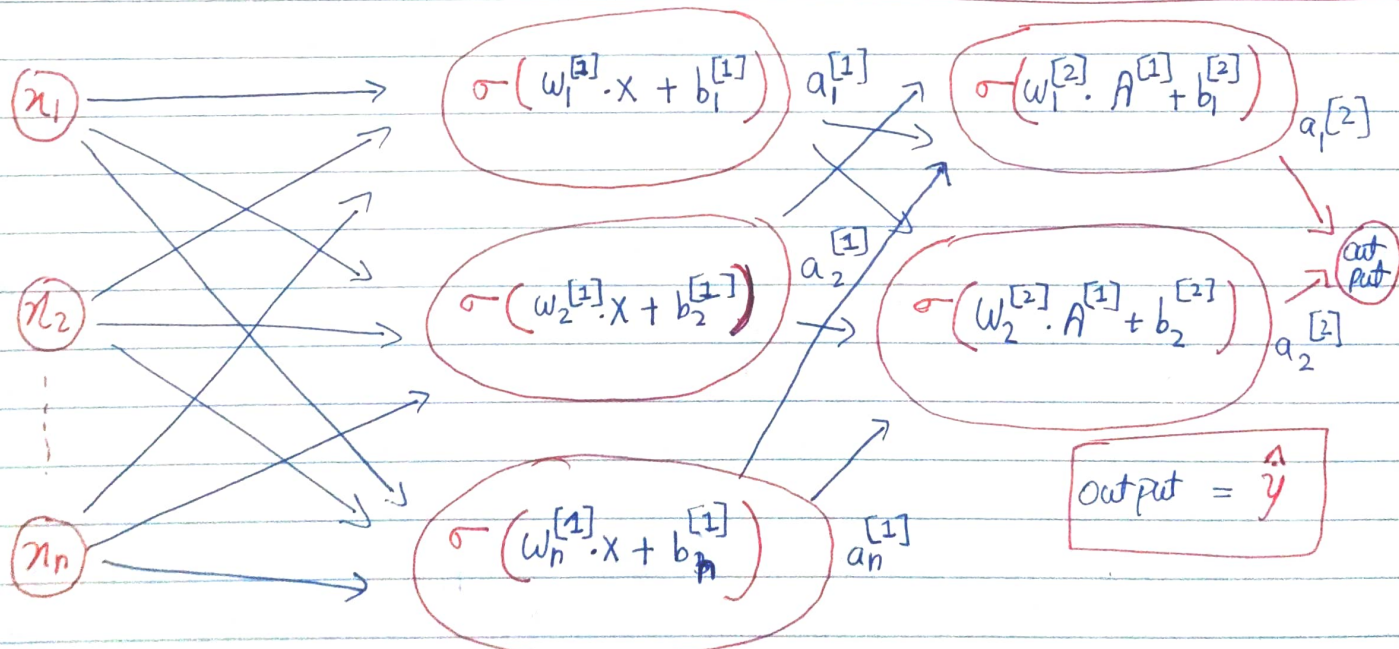
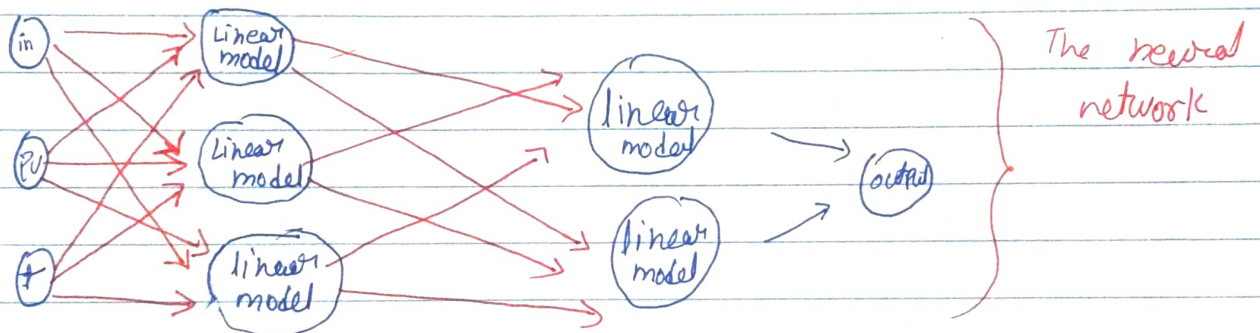
$$0a + 0b + 0c \neq 4$$

Contradictory or no solutions

The Neural Network

we have been doing linear regression, but what if the relation between the input & output is non-linear, in that case we use neural network which helps us in representing nonlinear systems. $\therefore$

$\Rightarrow$ A neural network is just large collection of linear models



The explanation of this whatever domination looking thing is on the next page

Explanation :-

⇒ Here the inputs are $(x_1), (x_2) \dots (x_n)$

⇒ we send all the value of (x) to each individual neuron in the next layer.

⇒ In that neuron we do ~~linear regression~~ have a unique set of w (weight) value & b (bias) associated with it, when multiplying weights to the inputs & adding bias we get a lot of equations

⇒ Now we shove the $w \cdot x + b$ expression into an activation function that gives us a vector (a)

⇒ Each of the neuron in that layer does exactly the same thing but with a different set of weights & biases. & gives us a vector output

In the second layer the inputs are the (a) values from the first layer.

Here $A^{[1]}$ is the matrix containing all the (a) vector's from the previous layer.

Now, the same stuff happens all over again. that generates vector $a^{[2]}$ & $a_2^{[2]}$ & this happens till you finally get the output (y) ⇒ This is also a vector.

Neural networks are really just a large collection of linear models that taken together can work to model highly non-linear systems.

In ML we use vectors, matrices & tensors to represent inputs & outputs of each layers.

Vectors & their properties

→ A vector is a simple array of numbers & has only one column, it can be depicted as ~~an~~ an arrow in the higher ~~dimensional~~ (plane & space)

→ Components of vector

→ Magnitude $\therefore$ its size

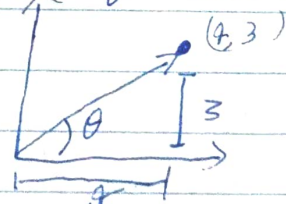
→ Direction

Norms → helps finding size of a vector.

L-1 norm = $|a, b|_1 = |a| + |b|$
(sum of absolute values of all the components)

L-2 norm = $|a, b|_2 = \sqrt{a^2 + b^2}$
we generally use L-2 norm. (The square root of sum of all the square components of the vector.)

Direction of a vector.



⇒ The direction of a vector can also be deduced from its coordinates.
 $\tan(\theta) = \frac{3}{4}$, $\theta = \arctan(3/4) = 0.64$ radian

Vector notation

Row vector

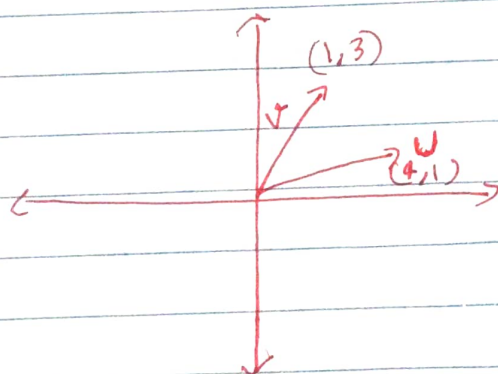
Column vector

$$x = (x_1 \ x_2 \ \dots \ x_n)$$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Vector operations

Sum of Vectors



$$u + v = (4+1, 1+3) = (5, 4)$$

$$u - v = (4-1, 1-3) = (3, -2)$$

General definition: Sum & difference

$$x = (x_1 \ x_2 \ \dots \ x_n)$$

$$y = (y_1 \ y_2 \ \dots \ y_n)$$

Sum

difference

$$x + y = (x_1 + y_1 \ x_2 + y_2 \ \dots \ x_n + y_n)$$

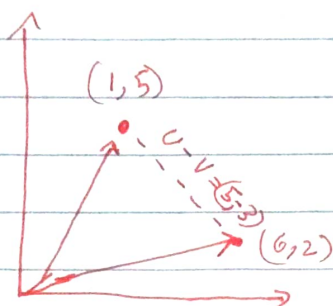
$$x - y = (x_1 - y_1 \ x_2 - y_2 \ \dots \ x_n - y_n)$$

Sum component by component

difference component by component.

⇒ To find the distance between two vectors just subtract them.

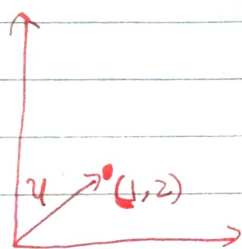
ex ⇒



L1-distance $|U-V|_1 = |5| + |3| = 8$

L2-distance $|U-V|_2 = \sqrt{5^2 + 3^2} = 5.83$

Multiplying a vector by a scalar



$$U = (1, 2)$$

$$\lambda = 3$$

$$\lambda U = (3, 6)$$

done in its ⁿ element wise multiplication.

Vector Dot product

quantities

Prices

Total price

2 apples

apples \$3

4 bananas

bananas \$5

1 cherry

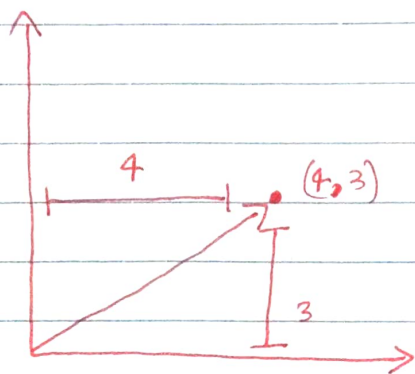
cherries \$2

$$\begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} \begin{matrix} \xrightarrow{\times} \\ \xrightarrow{\times} \\ \xrightarrow{\times} \end{matrix} \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} \begin{matrix} \rightarrow 2 \times 3 = 6 \\ \rightarrow 4 \times 5 = 20 \\ \rightarrow 2 \times 1 = 2 \end{matrix} \left. \vphantom{\begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}} \right\} \Rightarrow \boxed{28}$$

we can multiply each of our quantities & their respective prices to get the total price & that is dot product.

$$\begin{bmatrix} 2 & 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = 28$$

Norm of a vector using dot product



$$\sqrt{4^2 + 3^2} = \sqrt{25} = 5$$

$$\begin{bmatrix} 4 & 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 3 \end{bmatrix} = 25$$

$$L_2 \text{ norm} = \sqrt{\text{dot product } (u, u)}$$

$$\boxed{\|u\|_2 = \sqrt{\langle u, u \rangle}}$$

Vector Transpose

$$\begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = 28 \quad , \Rightarrow \quad \begin{bmatrix} 2 & 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = 28$$

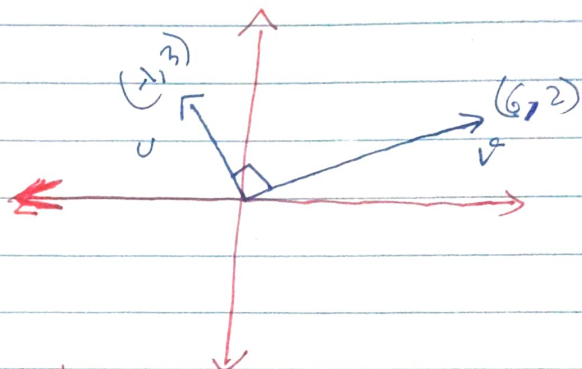
It is denoted by a superscript T.

$$\begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}^T = \begin{bmatrix} 2 & 4 & 1 \end{bmatrix}$$

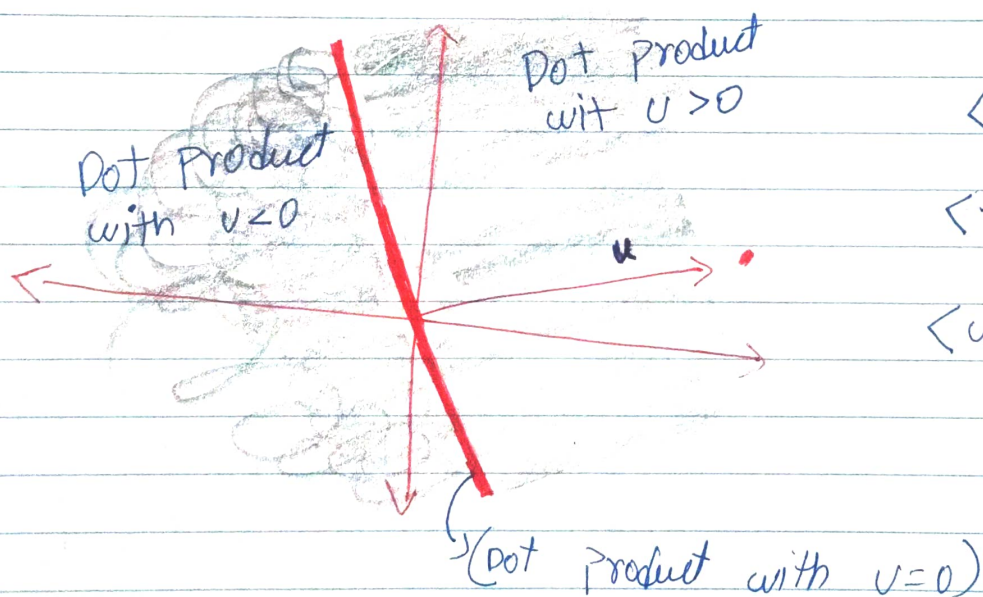
The geometric dot product

orthogonal or $\perp$ vector's have dot product as 0.

$$\begin{bmatrix} 6 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$



$$\langle u, v \rangle = |u| \cdot |v| \cos \theta$$



$$\langle u, v \rangle > 0$$

$$\langle u, v \rangle = 0$$

$$\langle u, v \rangle < 0$$

Multiplying Matrix by a vector

Equations as a dot product

$$\begin{aligned}a + b + c &= 10 \\a + 2b + c &= 15 \\a + b + 2c &= 12\end{aligned}$$

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 10 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 15 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 12 \end{bmatrix}$$

all of the dot products above
can be written as.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}_{3 \times 3} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix}_{3 \times 1} = \begin{bmatrix} 10 \\ 15 \\ 12 \end{bmatrix}$$

↳ The no of columns in vectors
& matrix should be equal if
we are doing multiplication
to no of.

The no of columns in matrix
should be equal to no of
rows in the vector

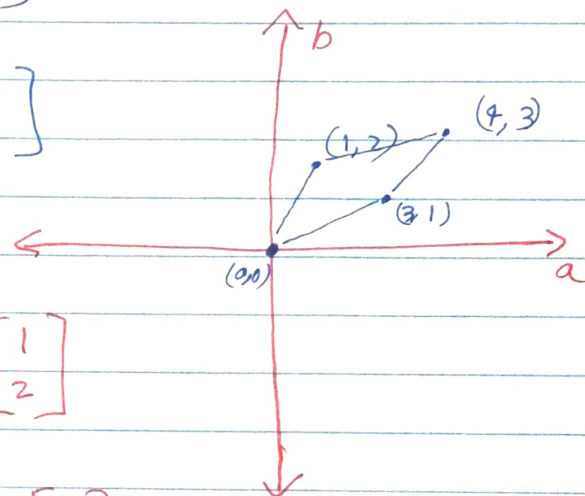
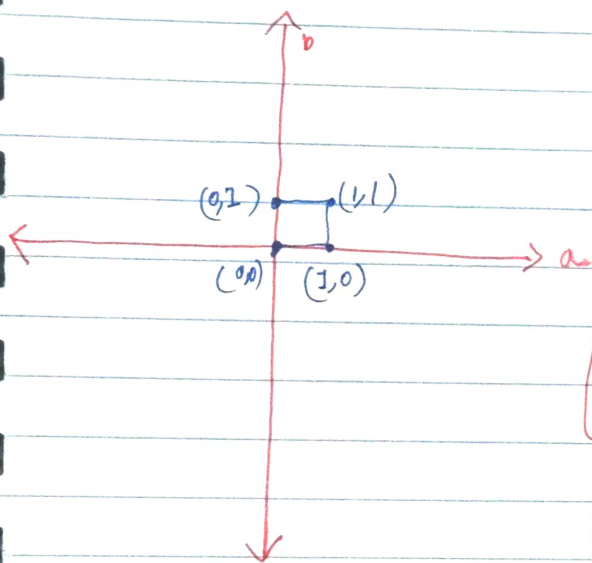
The no of columns
the vector should be equal to length of

Matrices as linear transformations

⇒ A linear transformation is a way to send each point in the plane into another point.
 "It helps in the representation of matrix"

$$\begin{bmatrix} a & b \\ 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$



$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

$$(0,0) \rightarrow (0,0) \rightarrow \text{cancel}$$

$$(1,0) \rightarrow (3,1) \rightarrow 3$$

$$(0,1) \rightarrow (1,2) \rightarrow 2$$

$$(4,3) \rightarrow (4,3) \rightarrow \text{cancel}$$

The square on the left & parallelogram on right are called basis.

Both the parallelogram tessellates the whole plane.

Linear transformations as matrices.

$$\begin{matrix} a & b \\ \begin{bmatrix} ? \\ ? \end{bmatrix} \end{matrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\begin{aligned} (0, 1) &\rightarrow (2, 3) \\ (1, 0) &\rightarrow (3, -1) \end{aligned}$$

$$\begin{matrix} a & b \\ \begin{bmatrix} ? \\ ? \end{bmatrix} \end{matrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

Here, the soln will be

$$\begin{bmatrix} 3 \\ -1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

& that's it

The identity matrix.

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

any matrix multiplied by Identity matrix remain same.

Matrix inverse

Inverse of a matrix is denoted by $[]^{-1}$

applying a little minus one on top.

$$\begin{bmatrix} 5 & 2 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Inverse of a matrix is a matrix when multiplied by its original form returns an identity matrix.

To solve it make the system of linear equations.

$$\Rightarrow 5a + 2c = 1$$

$$\Rightarrow 5b + 2d = 0$$

$$\Rightarrow a + 2c = 0$$

$$\Rightarrow b + 2d = 1$$

} solve these & you will get those values for inverse matrix

which matrix have inverse zero?

Only non-singular matrices will have inverse & singular matrices like redundant & contradictory will not have an inverse.

& non-singular matrices are invertible matrix & singular ones are non-invertible.

Neural Network's & matrices

let's take an example of spam email classifier.

Data :-

Spam	Lottery	win
Yes	1	1
Yes	2	1
No	0	0
Yes	0	2
No	0	1
No	1	0
Yes	2	2
Yes	2	0
Yes	1	2

from this data set we need to assign point's or scores to the no. lottery & win & a threshold, & if threshold point is crossed the email is spam

Scores
Lottery: 1 point's
win: 1 point's
Threshold: 1.5 point's

Examples

Lottery: 3 point's
win: 2 Points

"win, win the lottery!" :
7 point's { This sentence will have seven point's }.

The goal here is to find the best point & threshold

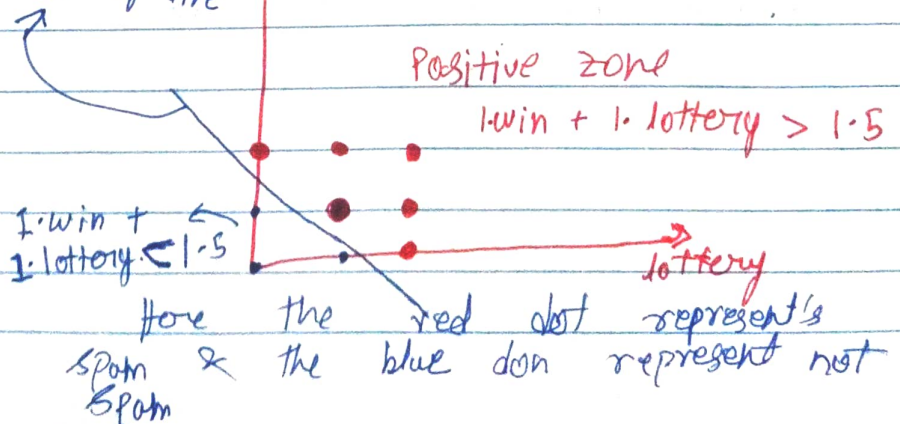
→ this is natural language processing

Score	>1.5?
2	Yes
3	Yes
0	No
2	Yes
1	No
1	No
4	Yes
2	Yes
3	Yes

→ graph win

$$1 \cdot \text{win} + 1 \cdot \text{lottery} = 1.5$$

eqn of line



⇒ The graph represents a linear classifier

⇒ A one layer neural network can be seen as a matrix product followed by a threshold check.

data →

Lottery	win	Spam
1	1	Yes
2	1	Yes
0	0	No
0	2	Yes
0	1	No
1	0	No
2	2	Yes
2	0	Yes
1	2	Yes

model

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

=

score

$$\begin{bmatrix} 2 \\ 3 \\ 0 \\ 2 \\ 1 \\ 1 \\ 4 \\ 2 \\ 3 \end{bmatrix}$$

(Spam or not)
~~score~~

> 1.5

$$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

⇒ The ~~data~~ dot product between data & model gives us the ^{vector} score, if the score greater than 1.5 then it's a spam

Threshold & bias

→ Here the check is

$$\Rightarrow 1 \cdot \text{win} + 1 \cdot \text{lottery} > \overbrace{1.5}^{\text{Threshold}}$$

$$\Rightarrow 1 \cdot \text{win} + 1 \cdot \text{lottery} - \overbrace{1.5}^{\text{bias}} > 0$$

→ The way to include this into matrix multiplication is that to add a bias ~~loop~~ column in the matrix with entries one & to add another row to the model with the bias (The no which is bias).

⇒ This is how it will look like.

Spam	lottery	win	Bias
Yes	1	1	1
Yes	2	1	1
No	0	0	1
Yes	0	2	1
No	0	1	1
No	1	0	1
Yes	2	2	1
Yes	2	0	1
Yes	1	2	1

model

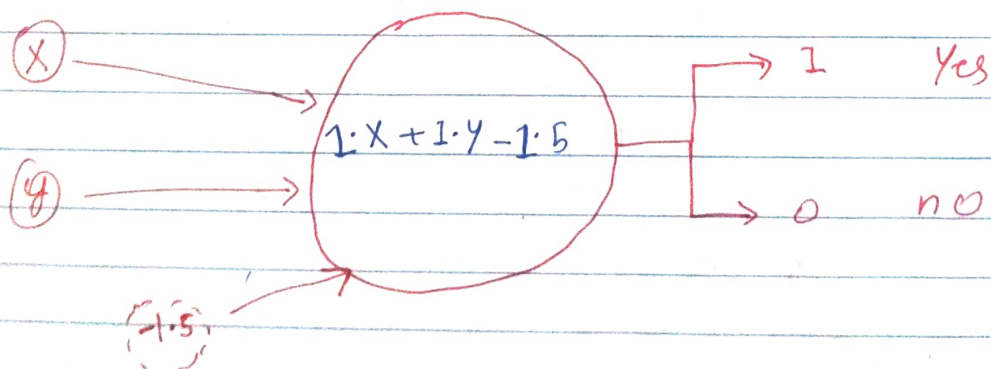
$$\begin{bmatrix} 1 \\ 1 \\ -1.5 \end{bmatrix}$$

Now, insted of checking if the bias is greater than 1.5 we only have to check, whether the score is +ve or -ve, if +ve then spam if -ve then not spam.

Perceptron

→ It is a one layer neural network

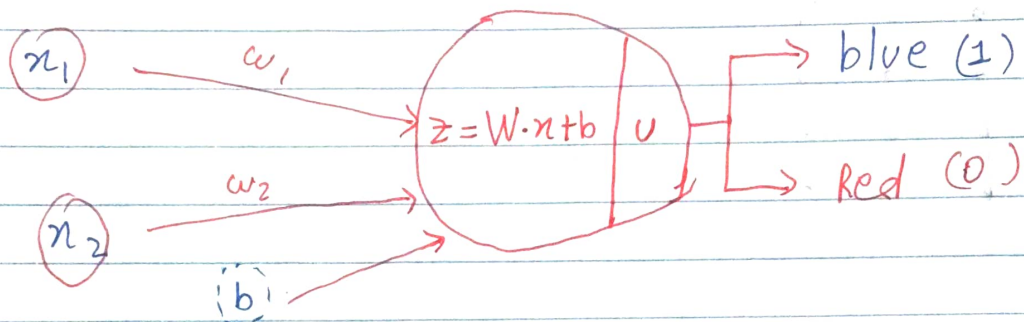
Here's a graphical way to represent previous perceptron



Here, X & Y are variable inputs & the one ("1")

multiplied to them is weight, 1.5 is a bias.
After the calculation, activation function checks whether the output is Yes or No.

This is the default graphical representation of single layer perceptron

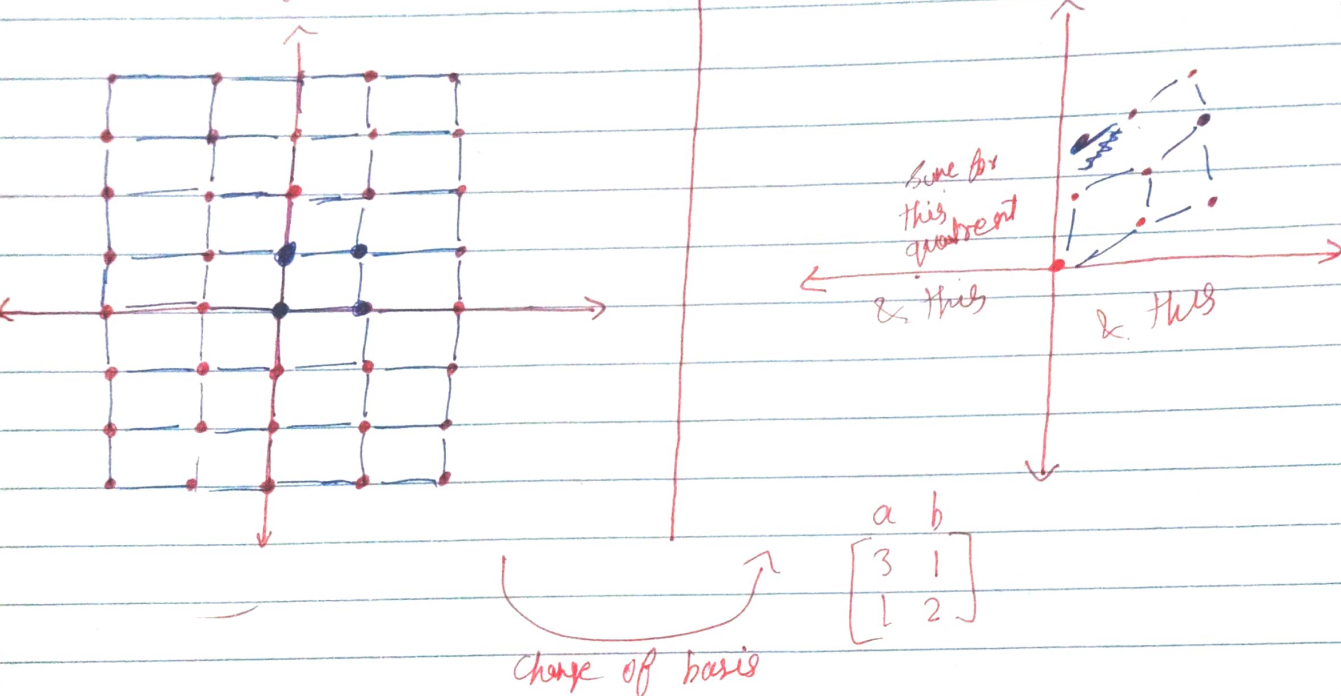


⇒ The output of the nodes get passed to activation fn, that generates the output of perceptron.

Here → x is a vector of input
 W is a vector of weights
 b is a bias

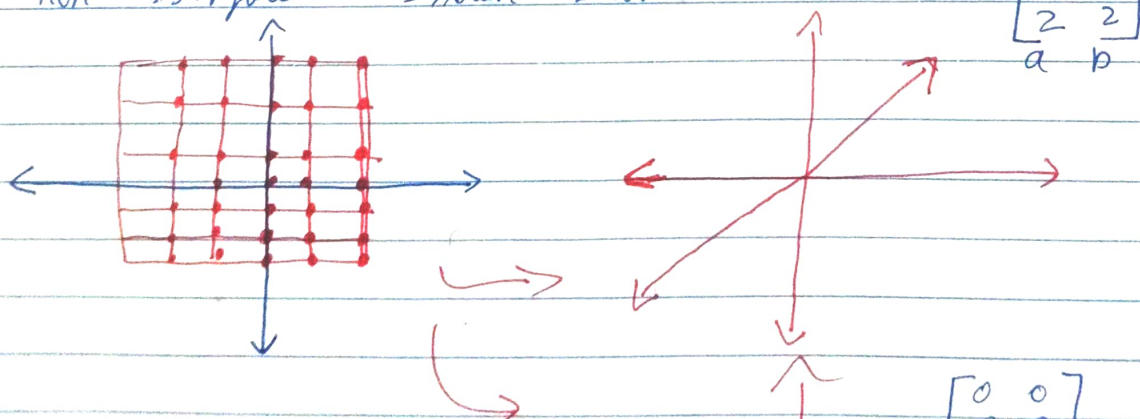
& U is an activation function, which gives 0 & 1
& the output label to the output.

Singularity & rank of linear transformations

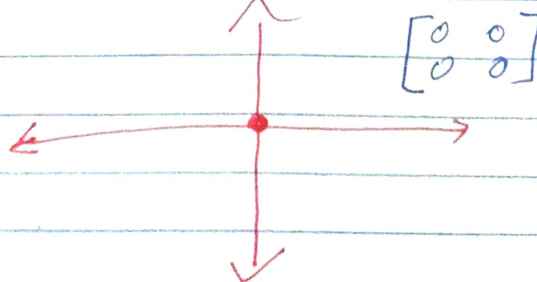


① $\Rightarrow$ while in linear transformation, when a matrix sends the whole plane to the whole plane, then it's singular. (like shown above)

② $\Rightarrow$ when a matrix sends a plane to a line it's non-singular. Shown below.



③ $\Rightarrow$ when a matrix sends a plane to dot it's also singular.



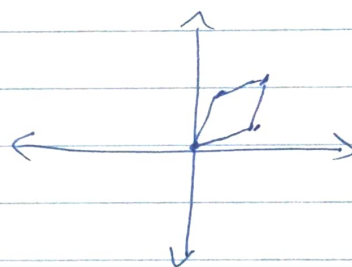
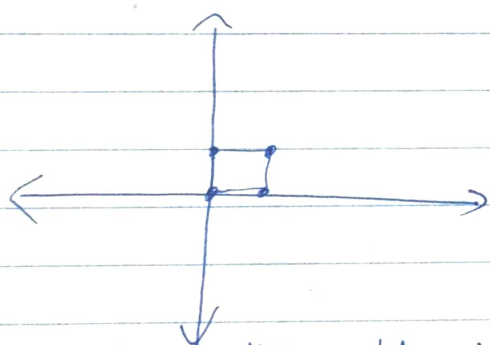
Notice that the dimension after the linear transform of ① is $\boxed{2}$, ~~② is $\boxed{1}$~~ & ~~③ is $\boxed{0}$~~ ② is $\boxed{1}$ & ③ is $\boxed{0}$. These are ~~its~~ ^{their} rank's $(2, 1, 0)$ respectively.

$\Rightarrow$ The dimension of the image of linear transformation is the rank.

Determinant as an Area

The determinants of a matrix $\therefore$

The area of the image of the fundamental basis formed by the unit square on the left.



$\rightarrow$ when the determinant is zero the matrix is non-singular.

Determinant of a product $\therefore$

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \times \begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ -3 & 3 \end{bmatrix}$$

$\det = 5$ $\det = 3$ $\det = 15$
 $(3 \cdot 2) - (1 \cdot 1)$ $(1 \cdot 1) - (-2 \cdot -2)$ $1 \cdot 3 - 4 \cdot (-3)$

$$\det(AB) = \det(A) \cdot \det(B)$$

The determinant of a matrix signifies that in linear transformation signifies that the area of the fundamental unit will get enlarged times the determinant (or any other area)

The transformation

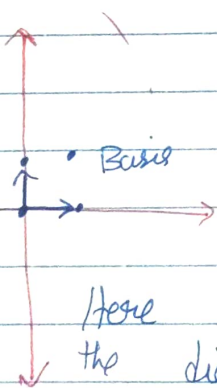
Determinant of inverse

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

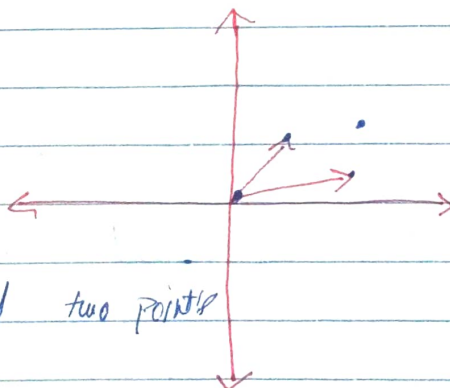
Bases in linear Algebra $\therefore$ A basis is a small set of vectors that you can use to build any vector in space by adding & scaling them.

Bases (Plural of "basis")

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$$



Here we only need two points the diagonal ones



Property $\therefore$ Every point in the space can be expressed as a combination of elements in the basis
 $\rightarrow$ It is linearly independent

What is not a basis?

$\rightarrow$ any two vectors having same direction & doesn't follow the property above

Span in linear algebra

→ The span of a set of vectors is the set of points that can be reached by walking in the direction of these vectors in any combination.

→ No. of elements in the basis is the dimension

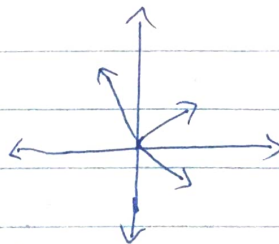
Linearly dependent & independent vectors.

Linearly independent vectors are the group of vectors in which none of the vectors in the group can be obtained as a linear combination of the others & the opposite for linearly dependent vectors

→ when vectors coincide → linearly dependent

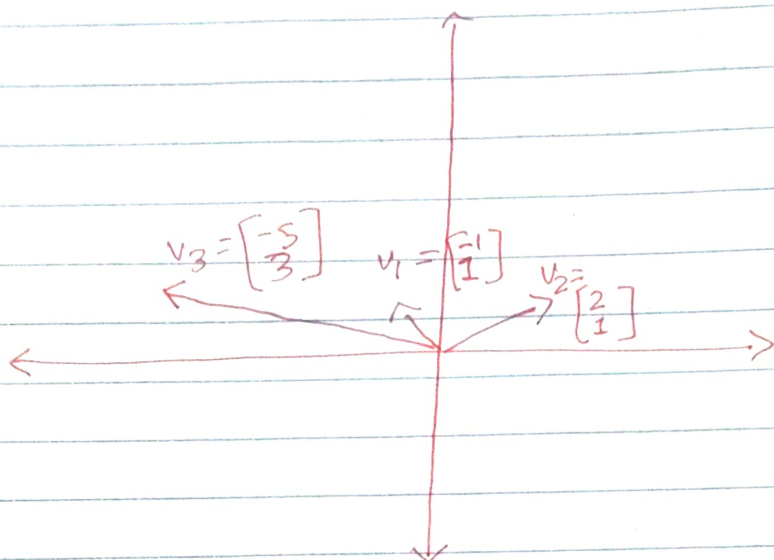
→ when vectors do not coincide → linearly independent

→ It turns out that any other vector you add can be made from the linear combination of other two hence linearly dependent



→ If you have more vectors than the dimension of the space you're trying to span, you will always have a linearly dependent group.

Check for linear dependence



we need to find constants α & β that satisfies eqn below
 $\Rightarrow \alpha v_1 + \beta v_2 = v_3$

$$\Rightarrow \alpha \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$$

this gives the eqn.

$\Rightarrow$

$$-\alpha + 2\beta = -5 \quad \text{--- (i)}$$

$$\alpha + \beta = 3 \quad \text{--- (ii)}$$

Solving (i) & (ii) we get

$$\Rightarrow \beta = \frac{-2}{3}, \quad \alpha = \frac{11}{3}$$

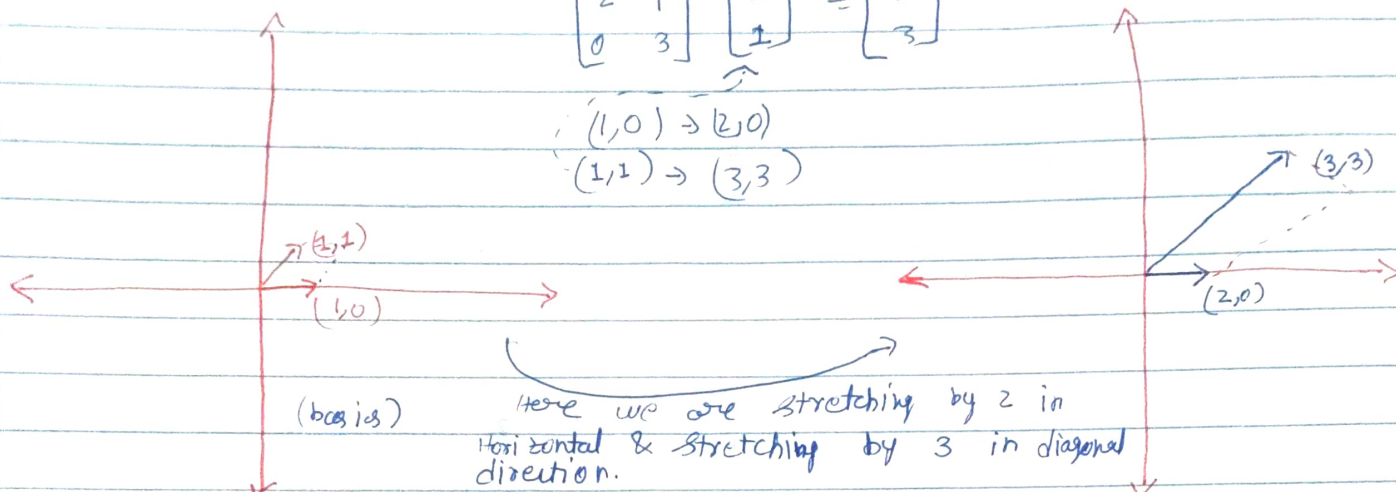
Since we were able to find the values of α & β , this means that v_3 is a linear combination of v_2 & v_1 hence it is linearly dependent.

Eigenbases

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$\begin{aligned} (1,0) &\rightarrow (2,0) \\ (1,1) &\rightarrow (3,3) \end{aligned}$$



⇒ Here, the sides of two hexagons are parallel to the corresponding one on the other basis. & the rest of the plane also follows.

⇒ This is a basis called eigenbasis; it only consists of two stretchings.

A vector that doesn't change their direction using linear transformation.

⇒ The two vectors in basis are called Eigen Vectors.

⇒ The ~~two~~ stretching factors are called Eigen values.

Eigen values & vectors :-

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$A V_1 = \lambda_1 V_1$$

$A \rightarrow$ matrix, $V_1 \rightarrow$ ^{eigen} Vector 1, $\lambda_1 \rightarrow$ Eigen value of given matrix of matrix given

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$A V_2 = \lambda_2 V_2$$

matrix multiplication } scalar multiplication
more compute } less compute

→ Along matrix eigen vectors, you can make large computation into smaller one.

ex →

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = ?$$

$\nearrow$ basis vector 1 $\nearrow$ basis vector 2 $\nearrow$ vector

→ The place where the vector goes in the new transformation

writing $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ in the linear combination of v_1 & v_2 vectors
 Solⁿ ⇒

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \left[-3 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right]$$

→ replacing $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ as the linear combination of v_1 & v_2

$$\Rightarrow -3 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} + 2 \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

→ we know the value

$$\Rightarrow -3 \begin{bmatrix} 2 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

of these two so all we need to do is replace them by the values

$$\Rightarrow \begin{bmatrix} -6 \\ 0 \end{bmatrix} + \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

- $Av = \lambda v$ for each eigenvector / eigenvalue
- Eigen ^{vectors} ~~values~~: direction of stretch
- Eigen ^{values} ~~vectors~~: how much stretch
- Eigen basis: the set of a matrix's eigenvectors, can be arranged as a matrix with one eigenvector in each column.
- Same work 2 help characterize linear transformation

Calculating Eigen Vectors & Eigen Values

→ The matrix must be square matrix, because we can find det of square matrix only.

⇒ $\det(A - \lambda I)$, → will give you characteristic polynomial & solving that will give you the values of lambda (λ) which is eigen values

ex →

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \quad \text{finding eigen values}$$

$$\Rightarrow \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} =$$

$$\Rightarrow \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Rightarrow \det \begin{bmatrix} (2-\lambda) & 1 \\ 0 & (3-\lambda) \end{bmatrix}$$

$$\Rightarrow (2-\lambda)(3-\lambda) - 1 \cdot 0 = 0$$

$$\boxed{\lambda = 2} \\ \boxed{\lambda = 3} \quad \text{eigen values.}$$

finding eigen vectors

for value $\lambda = 2$

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2 \begin{bmatrix} x \\ y \end{bmatrix}$$

$$2x + y = 2x \\ 0x + 3y = 2y$$

$$\begin{bmatrix} x=1 \\ y=0 \end{bmatrix}$$

~~1~~

for value $\lambda = 3$

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 3 \begin{bmatrix} x \\ y \end{bmatrix}$$

$$2x + y = 3x \\ 0x + 3y = 3y$$

$$\begin{bmatrix} x=1 \\ y=1 \end{bmatrix}$$

~~1~~

⇒ No of Eigen Vector's

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Eigen Values λ_1, λ_2

if $\lambda_1 \neq \lambda_2 \rightarrow 2$ eigen vector's
(2 different directions)

if $\lambda_1 = \lambda_2 \rightarrow 1$ eigen vector
(1 direction)
2 eigen vector
(2 different directions)

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

$\lambda_1, \lambda_2, \lambda_3$

if $\lambda_1 \neq \lambda_2 \neq \lambda_3 \rightarrow 3$ eigen vector's
(3 different directions)

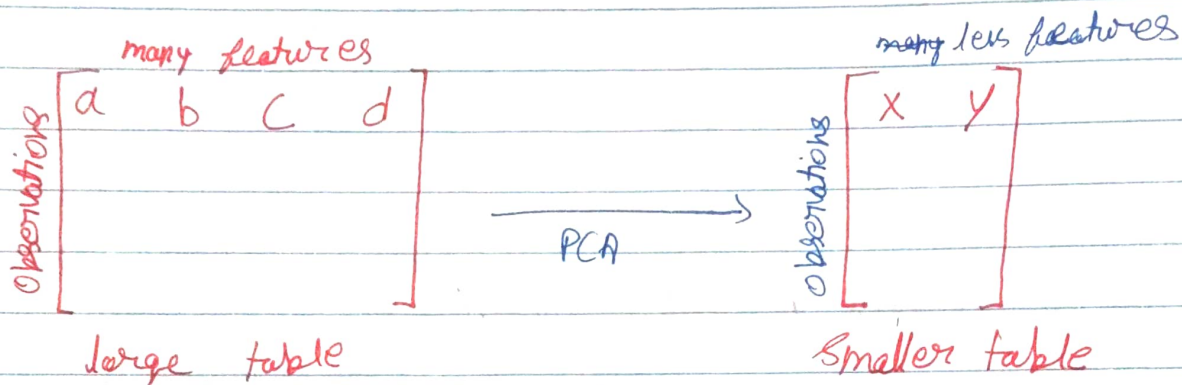
if $\lambda_1 = \lambda_2 \neq \lambda_3 \rightarrow 2$ eigen vector's
(2 diff directions)
↓ 3 eigen vector's
(3 different directions)

if $\lambda_1 = \lambda_2 = \lambda_3 \rightarrow 1$ eigen vector
(1 direction)

↓ 2 eigen vector
(2 different directions)
↓ 3 eigen vector's
(3 different directions)

Dimensionality Reduction & Projection

- Reduce dimensions or simply columns of dataset while preserving as much information as possible.



→ Same no of row's but less column's

→ Leads to smaller datasets

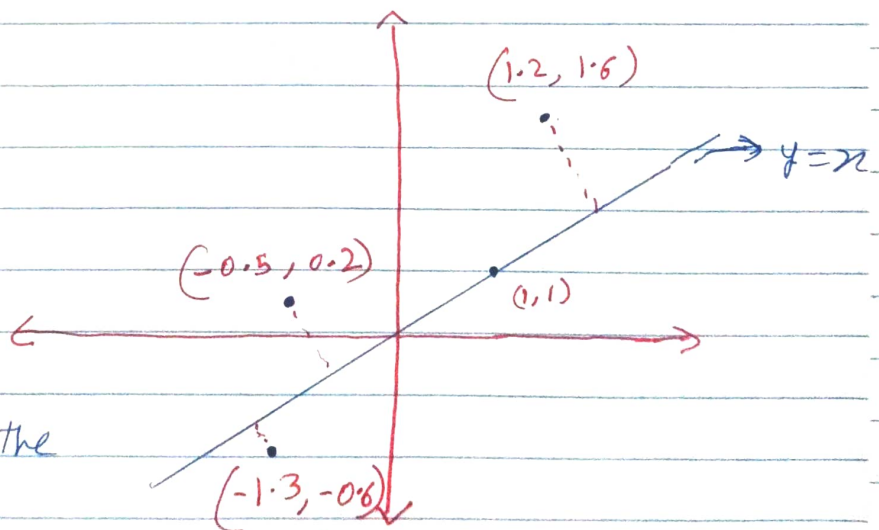
→ Easier to visualize

Projections

→ The idea is to move your ~~data~~ ^{vector} points into a vector ~~space~~ space with fewer dimensions.

Ex →

x	y
1.0	1.0
1.2	1.6
-0.5	0.2
-1.3	-0.6



Now moving your data on the line $y=x$.

Now, all you gotta do is perpendicularly project the data points onto the line & measure the distance from origin.

$$\begin{array}{cc}
 x & y \\
 \begin{bmatrix} 1.0 & 1.0 \\ 1.2 & 1.6 \\ -0.5 & 0.2 \\ -1.3 & -0.6 \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1}{\sqrt{2}} = \begin{bmatrix} \frac{(1+1)}{\sqrt{2}} \\ \frac{(1.2+1.6)}{\sqrt{2}} \\ \frac{(-0.5+0.2)}{\sqrt{2}} \\ \frac{(-1.3+(-0.6))}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1.4142 \\ 1.9799 \\ -0.2121 \\ -1.344 \end{bmatrix}
 \end{array}$$

To project a matrix A onto a vector V

$$\Rightarrow A_p = A \cdot \frac{V}{\|V\|_2}$$

Projected matrix Norm 2 of vector V

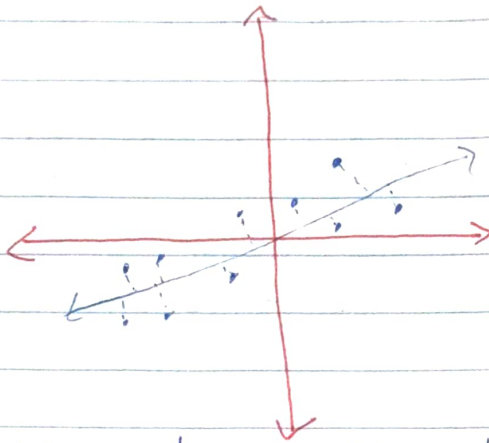
$\Rightarrow$ To project a matrix on multiple vectors

$$A_p = A \begin{bmatrix} \frac{v_1}{\|v_1\|_2} & \frac{v_2}{\|v_2\|_2} \end{bmatrix} \rightarrow \text{let's call this matrix "V"}$$

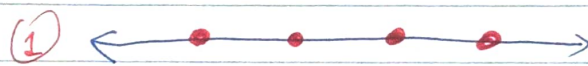
$$\boxed{A_p = AV}$$

Ans

Motivating PCA



→ Here we should choose the line where data is most spread out, because spreaded out data has preserved data the most



Here line 1 has most spread out data so we will choose it for PCA

Variance & Co-Variance

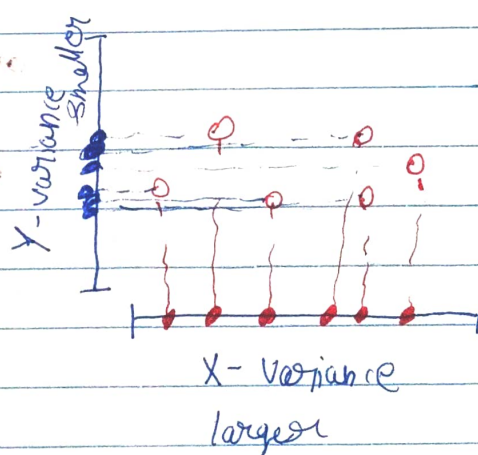
Mean: - The average of data

$$\text{Mean}(\bar{x}) = \frac{1}{n} \sum_{i=1}^n x_i$$

Variance: -

→ Variance ~~spread out~~ spread out when variance increases it indicates the data is ~~more~~ spreading out

→ When variance decreases the data becomes more compact.



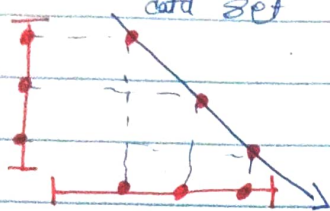
$$\text{Variance}(\bar{x}) = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \text{mean}(\bar{x}))^2$$

$$\text{Var}(\bar{x}) = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \mu_x)^2$$

it is average squared distance from the mean

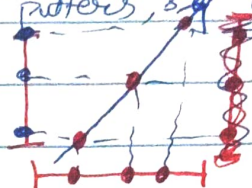
Problem with variance

⇒ Now these two data set will have same variance but these dataset show case different patterns, so what should we do?



(1)

Solution:- Covariance



(2)

Here as "x" values increase "y" values will decrease

"This has negative covariance"

Here as "x" values increase "y" values increase

"This dataset has +ve covariance"

Covariance

$$\text{cov}(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)$$

Here, $n \Rightarrow$ no of dataset

$x_i \Rightarrow$ The specific value of x from data

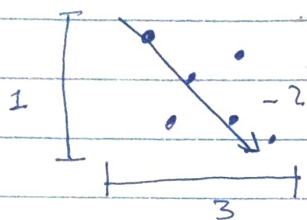
$\mu_x \Rightarrow$ The mean along x

$y_i \Rightarrow$ The specific value of y from data

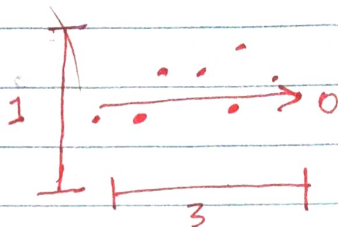
$\mu_y \Rightarrow$ The mean along y .

⇒ Here we are measuring the direction of the relationship btw two variables. it can be , negative trend if covariance is -ve , +ve trend if covariance is +ve & flat trend if very small covariance or no relationship

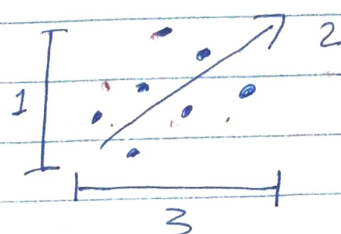
Co-variance Matrix



$$\begin{bmatrix} 3 & -2 \\ -2 & 1 \end{bmatrix}$$

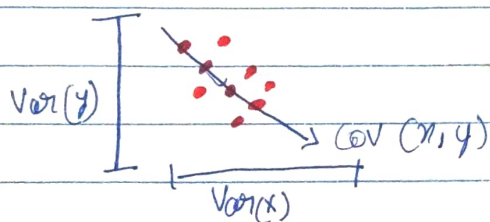


$$\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$$

These above are called Co-variance matrix.



Covariance matrix

$$C = \begin{matrix} & \begin{matrix} X & Y \end{matrix} \\ \begin{matrix} X \\ Y \end{matrix} & \begin{bmatrix} \text{Var}(x) & \text{Cov}(x, y) \\ \text{Cov}(y, x) & \text{Var}(y) \end{bmatrix} \end{matrix}$$

$$C = \begin{matrix} X & Y \\ \begin{bmatrix} \text{Cov}(x, x) & \text{Cov}(x, y) \\ \text{Cov}(y, x) & \text{Cov}(y, y) \end{bmatrix} \end{matrix}$$

(because $\text{var}(x) = \text{Cov}(x, x) \rightarrow$ just gauge at their formulas you will know.)

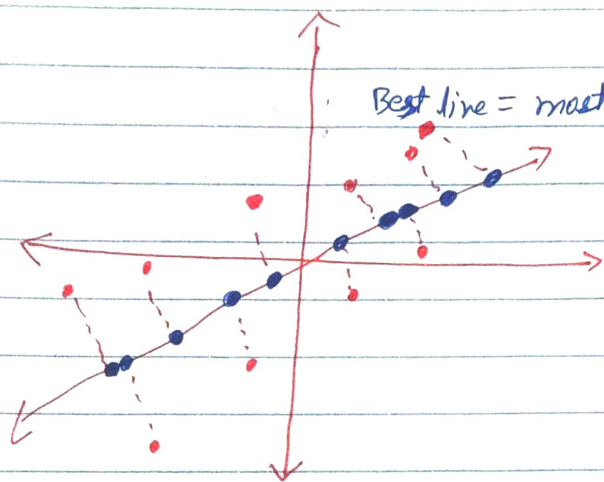
ex in matrix form.

$$A = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix}$$

$$U = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}$$

$$C = \frac{1}{n-1} (A - U)(A - U)^T \Rightarrow$$

Principal Component Analysis (PCA)



Projections
Eigen values / Eigen vector's
Covariance matrix

PCA

⇒ finding covariance $C = \begin{bmatrix} 3 & 4 \\ 4 & 3 \end{bmatrix}$

⇒ find the eigenvalues & eigen vector's of covariance matrix

⇒ The best line to project our data on will be the line made by one of the eigen vector's which has greater value of eigen value from the other one & will have greater value of variance

⇒ Now project the data on the line that goes through the chosen eigen vector

⇒ In the matrix's that has larger dimensions just pick up the greatest ^{eigen} values of eigen vector's, if you want the matrix to be reduced to two columns pick the top two ^{eigen} vector's of the matrix

⇒ Then use formula

$$X_{PCA} = \underbrace{(X - \mu)}_{\text{The vector chosen by us.}} V$$

↳ centered data, moves mean to the origin

Discrete Dynamical Systems

This is a system where we multiply a matrix by some vector & then we again multiply the matrix by the first output & keep repeating the process.

$$Ax = x_1$$

$$Ax_1 = x_2, \quad Ax_2 = x_3 \dots$$

	Sunny	Cloudy	rainy
Sunny	0.80	0.45	0.30
Cloudy	0.15	0.35	0.40
rainy	0.05	0.20	0.30

↳ reading this --- let's say it's sunny today then probability of tomorrow being sunny is 0.80, cloudy is 0.15, rainy is 0.05.

Here in this matrix, All values are positive & columns add to 1. This is a property of Markov matrix.

⇒ Markov matrix allow us to infer the probability of how your system will evolve.

let's say today is cloudy, then we can represent it like

	S	C	R
S	0.80	0.45	0.30
C	0.15	0.35	0.40
R	0.05	0.20	0.30

weather today, it is cloudy
 ↳ So every probability is 0 & cloudy is 1.

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$



state vector (x_0)
 ↳ It represents state of the system

$$\begin{matrix} \rightarrow \text{matrix} & \rightarrow v_0 \end{matrix}$$

$$\begin{bmatrix} 0.80 & 0.45 & 0.30 \\ 0.15 & 0.35 & 0.40 \\ 0.05 & 0.20 & 0.30 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.35 \\ 0.20 \end{bmatrix}$$

$\hookrightarrow x_1$

$$\text{matrix} \cdot x_1 = \begin{bmatrix} 0.5575 \\ 0.27 \\ 0.1525 \end{bmatrix}$$

$$\text{matrix} \cdot x_2 = \begin{bmatrix} 0.6293 \\ 0.2421 \\ 0.1286 \end{bmatrix} \quad \hookrightarrow x_2$$

$$\text{matrix} \cdot x_\infty = \begin{bmatrix} 0.6665 \\ 0.2223 \\ 0.1112 \end{bmatrix}$$

$\downarrow$

$$\boxed{\text{matrix} \cdot x_n = x_{n+1}}$$

$\hookrightarrow$ at infinity

$$\text{matrix} \cdot x_n = x_{n+1}$$

at infinity

$$\boxed{x_n = x_{n+1}}$$

Hence we will get to the eigen vector with eigen value 1.

$$\begin{bmatrix} 0.80 & 0.45 & 0.30 \\ 0.15 & 0.35 & 0.40 \\ 0.05 & 0.20 & 0.30 \end{bmatrix} \begin{bmatrix} 0.6665 \\ 0.2223 \\ 0.1112 \end{bmatrix}$$

$\rightarrow$ It doesn't matter what your initial vector was after computing till ∞ you will end up having this vector

$\hookrightarrow$ Transition matrix (P) $\hookrightarrow$ equilibrium vector (x_∞)

"It gives you a long run probability that on a given day it will be sunny, cloudy or rainy"