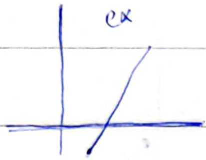
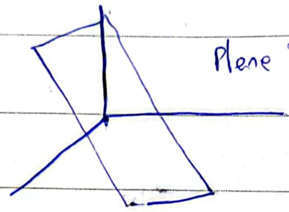
linear Algebra is the most usefull & widespread math Field in ML

Systems of linear equations → linear Regression → supervised ML (Already have data (input & output). goal: discover relation between inp & outp

$$\overset{\text{weight}}{w}\,\overset{\text{feature}}{x} + \overset{\text{bias}}{b}: \quad y = wx + b \qquad y = w_1 x_1 + w_2 x_2$$

ex





Plane in 3d

$$w_1 x_1^{(1)} + \dots + w_n x_n^{(1)} + b = y^{(1)}$$
$$w_1 x_1^{(2)} + \dots + w_n x_n^{(2)} + b = y^{(2)}$$
$$\vdots$$
$$w_1 x_1^{(m)} + \dots + w_n x_n^{(m)} + b = y^{(m)}$$

$$W \cdot X + b = y$$

★ x_i in each row is dif from another row

★ w_i & b are same in every row

↳ $[w_1 \ w_2 \dots w_n]$ Vector

$$\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_n^{(2)} \\ \vdots & \vdots & & \vdots \\ x_1^{(m)} & x_2^{(m)} & \dots & x_n^{(m)} \end{bmatrix}$$

Matrix

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

Vector

Systems of Sentences

System 1	System 2	System 3
The dog is black	The dog is black	The dog is black
The Cat is orange	The dog is black	The dog is white
Complete	redundant	Contradictory
non-singular	Singular	singular

linear equation:

$a + b = 10$

$2a + 3b = 15$

Non-linear

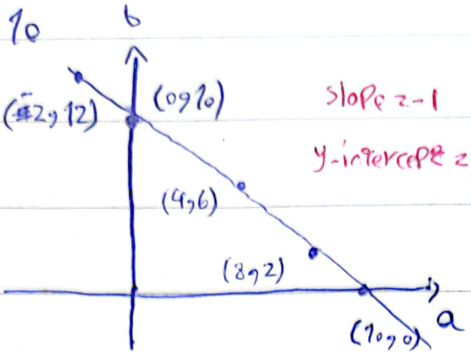
$a^2 + b^2 = 10$, $\sin(a) + b^2 = 15$

$2^a - 3^b = 0$

$ab^2 + \frac{b}{a} - \frac{3}{b} - \log(c) = 4^a$

Linear equation → line

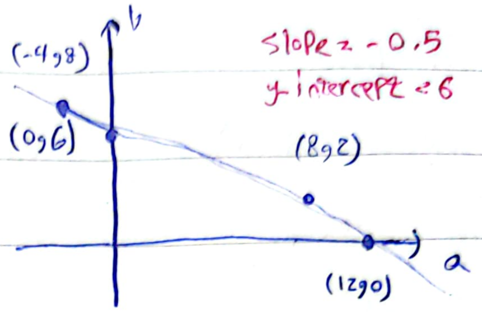
$a+b=10$



slope = -1

y-intercept = 10

$a+2b=12$



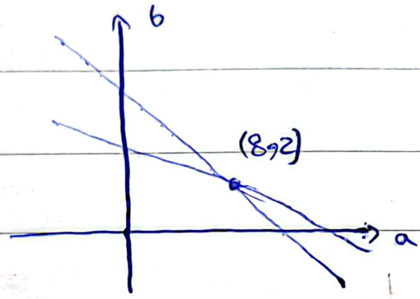
slope = -0.5

y-intercept = 6

$a+b=10$

$a+2b=12$

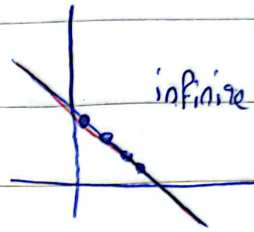
Non-singular



$a+b=10$

$2a+2b=20$

singular



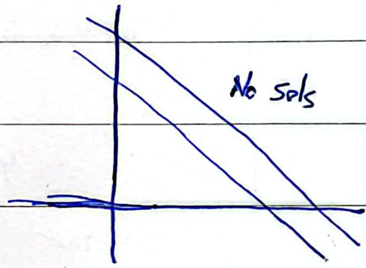
infinite sols

Redundant

$a+b=10$

$2a+2b=24$

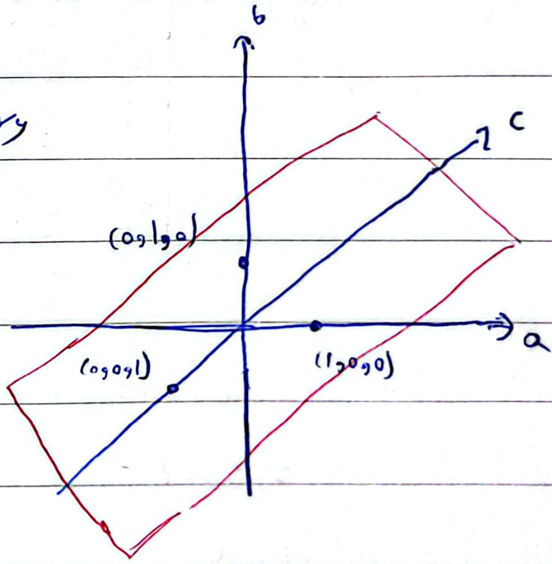
singular



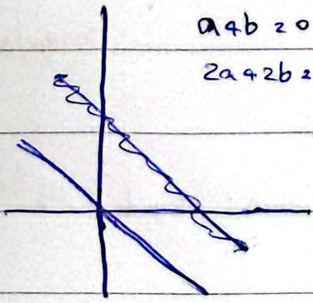
No sols

contradictory

$a+b+c=1$

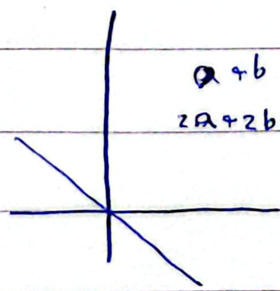


★ we can set the constant to 0 so that singular system would be one type which is redundant (contradictory systems turns to redundant)



$a+b=0$ (10)

$2a+2b=0$ (20)



$a+b=0$ (10)

$2a+2b=0$ (24)

singular

★ the important part:

Singularity & Non-singula

Systems of equations as matrices

$a + b = 0$
$a + 2b = 0$
$\longrightarrow$
$$\begin{array}{cc} a & b \\ \end{array}$$
$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

Non-singular system — Non-singular matrix

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

singular ~~matrix~~ system — singular matrix

★ Constant don't matter for singularity

Linear dependence between rows

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$$ $2\,\text{Row}1 = \text{Row}2$ → Singular
(linearly dependent)

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$ linearly independent → Non-singular

Ex

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$ Row1 + Row2 = Row3

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix}$$ Row1 + Row2 = Row3

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 1 & 3 \end{bmatrix}$$ Avg(Row1 + Row3) = Row2

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

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 3 & 2 & 3 \end{bmatrix}$$ 3Row1 + 2Row2 = Row3

$$\begin{bmatrix} 1 & 2 & 5 \\ 0 & 3 & -2 \\ 2 & 4 & 10 \end{bmatrix}$$ 2Row1 = Row3

★ Determinant of singular matrix is 0

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \quad 1\times 2 - 1\times 2 = 0 \qquad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

$$1(4-1) - 1(2-1) + 1(1-2) = 3 - 1 - 1 = 1$$

Row echelon form

★ on main diag, we have bunch of ones followed by perhaps a bunch of zeros. (we can have all 1's or all 0's)

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

★ below diag → everything i[s] zero

★ to right of 1's → any n[umber] is allowed

★ to the right of zero's → every thing must be zeros

For 2×2:

$$\begin{bmatrix} 1 & * \\ 0 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 1 & * \\ 0 & 0 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Row operations

1- Swiching rows:

$$\underset{\substack{\text{non-singular} \\ \det = 11}}{\begin{bmatrix} 5 & 1 \\ 4 & 3 \end{bmatrix}} \longrightarrow \underset{\substack{\text{non-singular} \\ \det' = -11}}{\begin{bmatrix} 4 & 3 \\ 5 & 1 \end{bmatrix}} \rightarrow \det' = -\det$$

2- Multiplying a row by a scalar:

$$\underset{\det = 11}{\begin{bmatrix} 5 & 1 \\ 4 & 3 \end{bmatrix}} \xrightarrow{\times 10} \underset{\det' = 110}{\begin{bmatrix} 50 & 10 \\ 4 & 3 \end{bmatrix}} \qquad \det' = 10\det$$

3- Adding a row to another row:

$$\underset{\det = 11}{\begin{bmatrix} 5 & 1 \\ 4 & 3 \end{bmatrix}} \longrightarrow \underset{\det' = 11}{\begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix}} \qquad \det' = \det$$

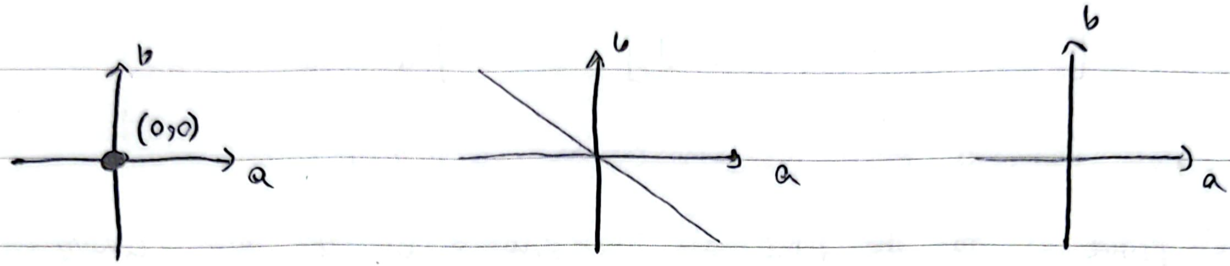
Rank of a system: the amount of information a system carries.

$a+b=0$, $a+2b=0$ $\begin{bmatrix}1 & 1\\ 1 & 2\end{bmatrix}$ Rank: 2

$a+b=0$, $2a+2b=0$ $\begin{bmatrix}1 & 1\\ 2 & 2\end{bmatrix}$ Rank: 1

$0a+0b=0$, $0a+0b=0$ $\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$ Rank: 0

Dimension of solution spaces:



0-d 1-d 2-d

★ Rank = 2 - (Dimension of solution space)

★ a matrix is non-singular if it has full rank.

How to get the row echelon form: ①

$$\begin{bmatrix}5 & 1\\ 4 & -3\end{bmatrix} \xrightarrow[\text{by leftmost coef}]{\text{Divide each row}} \begin{bmatrix}1 & 0.2\\ 1 & -0.75\end{bmatrix} \xrightarrow[\text{Row 2 = Row 2 - Row 1}]{\text{to get rid of}} \begin{bmatrix}1 & 0.2\\ 0 & -0.95\end{bmatrix}$$

$$\xrightarrow[\text{the leftmost non-zero coef}]{\text{Divide the 2nd Row by}} \overset{\text{ref ①}}{\begin{bmatrix}1 & 0.2\\ 0 & 1\end{bmatrix}}$$ Rank = 2

$$\begin{bmatrix}5 & 1\\ 10 & 2\end{bmatrix} \rightarrow \begin{bmatrix}1 & 0.2\\ 1 & 0.2\end{bmatrix} \rightarrow \overset{\text{① ref}}{\begin{bmatrix}1 & 0.2\\ 0 & 0\end{bmatrix}}$$

singular Rank = 1

$$\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix} \rightarrow \overset{\text{① ref}}{\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}}$$ Rank = 0

★ Rank = the number of ones in the diagonal of ref

$$\begin{bmatrix}1 & 1 & 2\\ 3 & -3 & -1\\ 2 & -1 & 6\end{bmatrix} \rightarrow \overset{\text{ref}}{\begin{bmatrix}1 & 1 & 2\\ 0 & -6 & -7\\ 0 & 0 & 6\end{bmatrix}}$$

$a+b+2c=12$

$3a-3b-c=3$

$2a-b+6c=24$

Row echelon form in general

Rank = 5

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

Rank = 3

$$\begin{bmatrix} 3 & \star & \star & \star & \star \\ 0 & 0 & 1 & \star & \star \\ 0 & 0 & 0 & -9 & \star \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- $\star$ Each row has a Pivot (left-most non-zero entry)
- $\star$ Every pivot is to the right of the pivots on the rows above
- $\star$ Rank of the matrix is the number of Pivots

Reduced REF :

$5a + b = 17$	$a + 0.2b = 3.4$	$a = 3$
$4a - 3b = 6$	$b = 2$	$b = 2$

REF

$$\begin{bmatrix} 5 & 1 \\ 4 & -3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0.2 \\ 0 & 1 \end{bmatrix} \xrightarrow[R_1 - R_2 \times 0.2]{R_1 = \cancel{R_2 \times 0.2}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Rules for reduced REF:

1- It has to be in REF
2- Each Pivot is a 1
3- Any number above a Pivot is 0
4- Rank = # Pivots

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Augmented matrix : Matrix of coefs + a column for constant values

Gaussian Elimination - Summary:

1- Create the augmented matrix
2- Get the matrix into R-REF
3- Complete back substitution
4- Stop if you encounter a row of 0s

? when reaching a row of 0s

☆ if constant in that row is zero → Infinitely many solutions

☆ if constant in that row is not zero → No solutions

Ex

$2a - b + c = 1$

$2a + 2b + 4c = -2$

$4a + b = -1$

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

$$\left[\begin{array}{ccc|c} 1 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 3 & 3 & -3 \\ 0 & 3 & -2 & 2 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 1 & -1 \\ 0 & 3 & -2 & 2 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & -1 & -1 \\ 0 & 0 & -5 & 5 \end{array}\right]$$

REF

$$\rightarrow \left[\begin{array}{ccc|c} 1 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & -1 \end{array}\right] \xrightarrow[\text{substitution}]{\text{Back}}$$

☆ start from bottom Rows use the pivot from each row to cancel the values in the cells above it

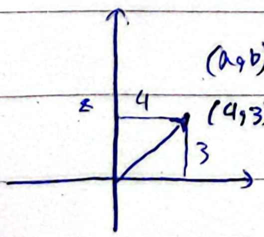
$$\xrightarrow[R1 - \frac{1}{2}R3]{R_2 - R3} \left[\begin{array}{ccc|c} 1 & -\frac{1}{2} & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{array}\right] \xrightarrow{R1 = R1 + \frac{1}{2}R2} \underbrace{\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{array}\right]}_{\text{identity matrix}}$$

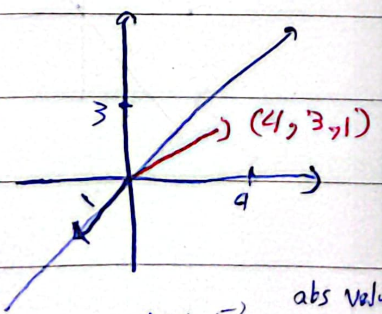
$a = 1$

$b = 0$

$c = -1$

Vector

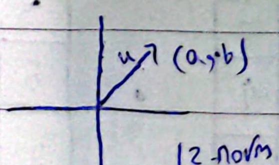




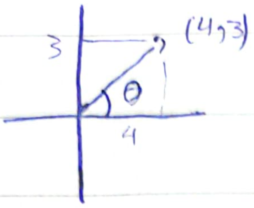
Norm:

$L1\text{-norm} = \|(a, b)\|_1 = |a| + |b|$ → abs value

$L2\text{-norm} = \|(a, b)\|_2 = \sqrt{a^2 + b^2} = u$



L2-norm ✓



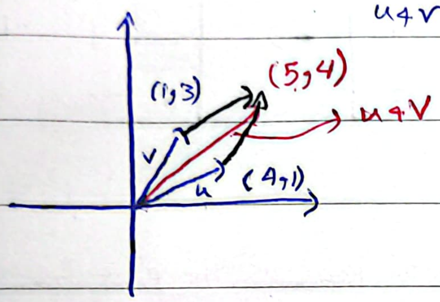
$\tan(\theta) = \frac{3}{4} \rightarrow \theta = \arctan\left(\frac{3}{4}\right) = 0.64 \text{ rad} = 36.87^\circ$

Vector notation

$x = (x_1, x_2 \cdots x_n)$ $\quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ or $[x_1 x_2 \cdots x_n]$ $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

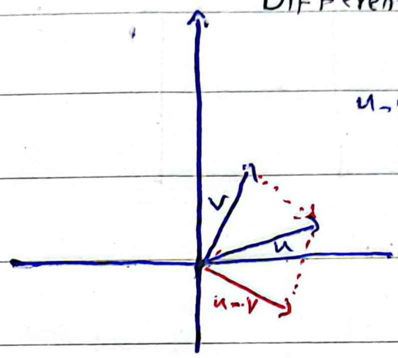
Vector Operations

Sum



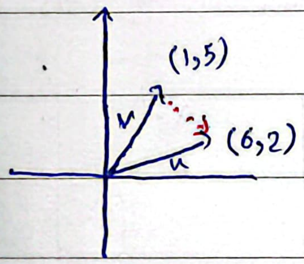
$u+v = (5,4)$

Difference



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

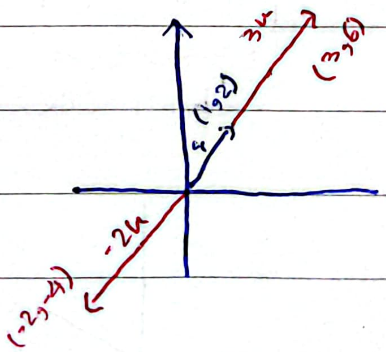
Distances



$L1 = |u-v|_1 = |5| + |-3| = 8$

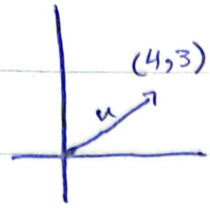
$L2 = |u-v|_2 = \sqrt{5^2+3^2} = 5.83$

Multiplying by a Scalar



The dot Product

$\begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = 2.3 + 4.5 + 1.2 = 28 \equiv \begin{bmatrix} 2 & 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}$

★ 

$|u|_2 = \sqrt{4^2+3^2} = 5 \equiv \sqrt{\begin{bmatrix}4\\3\end{bmatrix} \cdot \begin{bmatrix}4 & 3\end{bmatrix}}$

→ $L_2\text{-norm} = \sqrt{\text{dot Product}(u,u)}$ or $|u|_2 = \sqrt{\langle u,u \rangle}$ → d-Prod

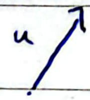
Matrix Transpose

$\begin{bmatrix}2\\4\\1\end{bmatrix}^T \rightarrow \begin{bmatrix}2 & 4 & 1\end{bmatrix}$ $\quad \begin{bmatrix}2 & 4 & 1\end{bmatrix}^T \longrightarrow \begin{bmatrix}2\\4\\1\end{bmatrix}$

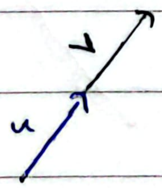
$\begin{bmatrix}2 & 5\\4 & 7\\1 & 3\end{bmatrix}^T \longrightarrow \begin{bmatrix}2 & 4 & 1\\5 & 7 & 3\end{bmatrix}$

3x2 2x3

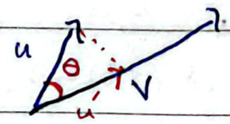
The dot Product in general



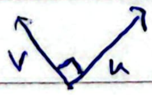
$\langle u,u \rangle = |u|^2$



$\langle u,v \rangle = |u| . |v|$



$\langle u,v \rangle = |u'| . |v| = |u| |v| \cos\theta$

★ → 

$\langle u,v \rangle = |u| |v| \cancel{\cos 90^\circ} = 0$

Equations as dot Product

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

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

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

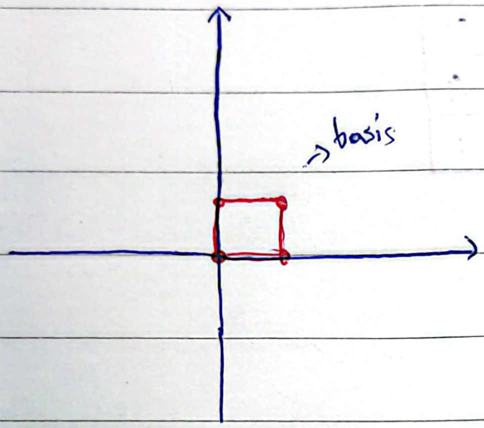
or →

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

3 x 3 — Vector length = 3 — length = 3

★ # Columns = length of vector

Matrices as linear Transformation : send each point in the plane into another point in the plane



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

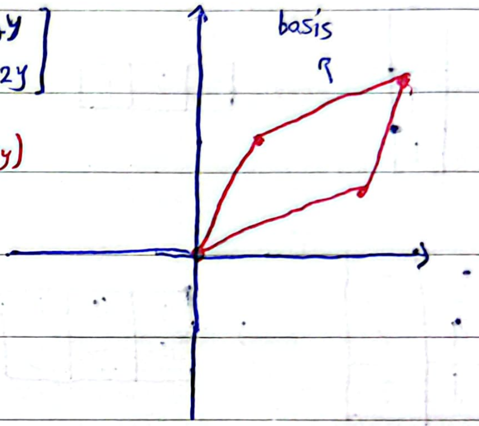
$(x, y) \rightarrow (3x+y, x+2y)$

$(0,0) \rightarrow (0,0)$

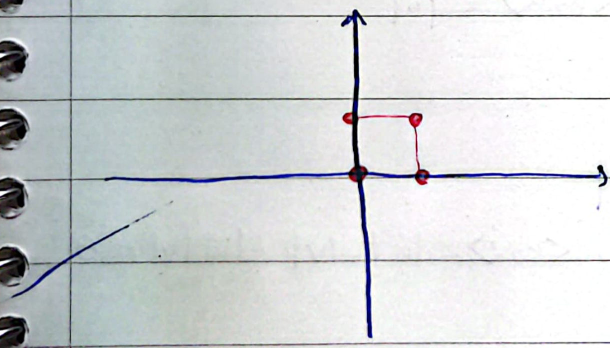
$(1,0) \rightarrow (3,1)$

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

$(1,1) \rightarrow (4,3)$



Linear Transformation as Matrices



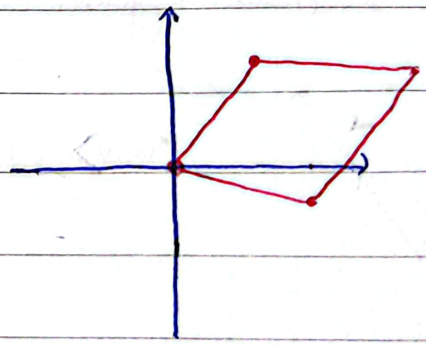
$$\begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$(0,0) \rightarrow (0,0)$

★ $(1,0) \rightarrow (3,-1)$ ①

★ $(0,1) \rightarrow (2,3)$ ②

$(1,1) \rightarrow (5,2)$



①

$$\begin{bmatrix} x & y \\ w & z \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} x \\ w \end{bmatrix} \xrightarrow{=} \begin{bmatrix} 3 \\ -1 \end{bmatrix} \rightarrow \begin{bmatrix} 3 & ? \\ -1 & ? \end{bmatrix}$$

②

$$\begin{bmatrix} 3 & y \\ -1 & z \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} y \\ z \end{bmatrix} \xrightarrow{=} \begin{bmatrix} 2 \\ 3 \end{bmatrix} \rightarrow \begin{bmatrix} 3 & 2 \\ -1 & 3 \end{bmatrix}$$

Multiplying Matrices

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{bmatrix}$$

★ Columns of first matrix must match rows of second

★ Result takes number of rows from first matrix

★ 〃 〃 〃 〃 columns 〃 second 〃

The identity matrix:

the matrix that when multiplied by any other matrix, it gives the same matrix.

$$\begin{bmatrix} 1 & 0 & \cdots & & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & 1 \cdots & 0 \\ 0 & 0 & 0 & 0 & 1 \cdots 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix}$$

Inverse of a matrix:

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{cases} \to 3a+c=1 & \to a=\frac{2}{5} \\ \to 3b+d=0 & \to b=-\frac{1}{5} \\ \to a+2c=0 & \to c=-\frac{1}{5} \\ \to b+2d=1 & \to d=\frac{3}{5} \end{cases}$$

★ Singular matrices never have an inverse → non-invertible [det = 0]

NLP!

We want to decide whether an email is spam or not → spam filter (classifier)

we have a spam dataset in which we pinpoint # 2 words 'Lottery' & 'win' that are quite deterministic for a spam.

Subject: Date:

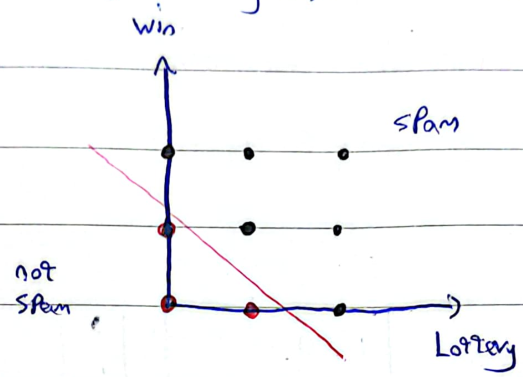
Spam	Lottery	Win
Y	1	1
Y	2	1
N	0	0
Y	0	2
N	0	1
N	1	0
Y	2	2
Y	2	0
Y	1	2

Table 1

→ we need to find x → score for Lottery and y → score for win and then calculate the whole sentence's score and compare it to threshold

if score > threshold → spam if not → not spam

→ Lottery → 1 win → 1 threshold → 1.5



★ One-layer NN is a matrix product followed by a threshold check

Table 1 $\cdot$ $\begin{array}{c} \text{Model} \\ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{array}$ =

Prod
2
3
0
2
1
1
4
2
3

→ Check > 1.5?

Threshold

1.win + 1.lottery > 1.5

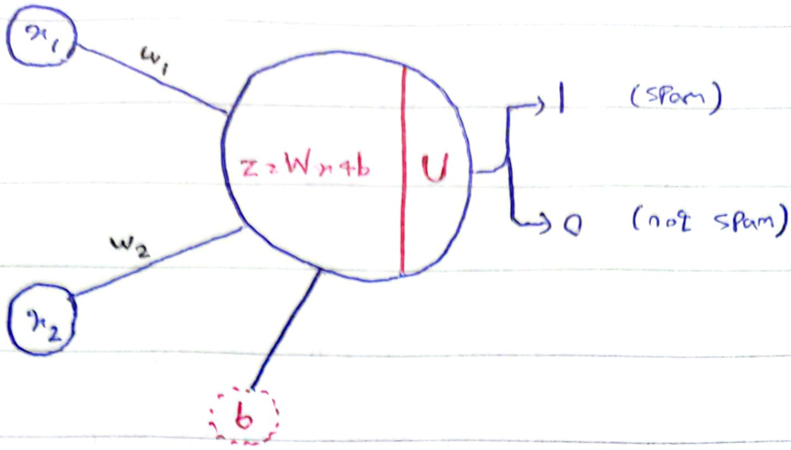
★ Another way

Lottery	win	Bias
1	1	1
2	1	1
0	0	1
0	2	1
0	1	1
1	0	1
2	2	1
2	0	1
1	2	1

$\cdot$ $\begin{array}{c} \text{Model} \\ \begin{bmatrix} 1 \\ 1 \\ -1.5 \end{bmatrix} \end{array}$ →

bias

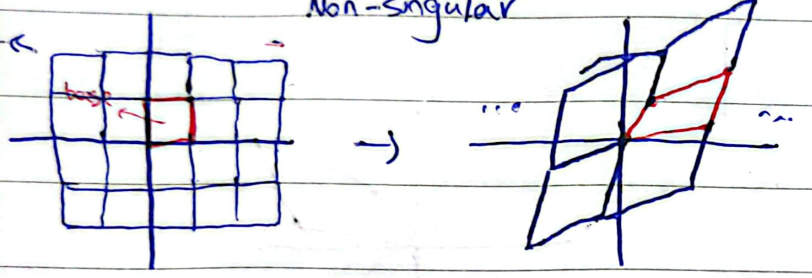
1.win + 1. Lottery − 1.5 > 0



Rank of linear Transformation - Singular / Non-singular

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

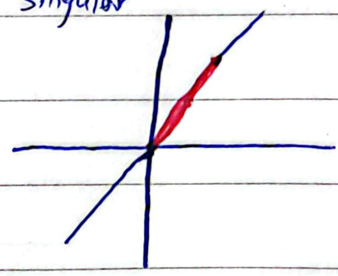
① Non-singular



Rank = Dimension = 2

Singular

M1 →



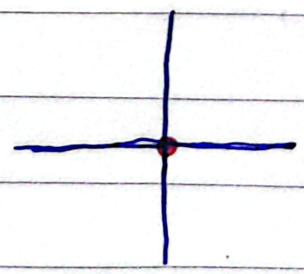
Rank = Dimension = 1

① : If the resulting points after multiplying it by a matrix cover the entire plane → Non-singular & vice-versa

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

singular

M1 →



Rank = Dimension = 0

Subject: Date:

Determinant as an area: the determinant of a matrix [it's better to say it's abs value] is the area of the img of the fundamental basis

☆ Singular transformations → det = 0 so area = 0

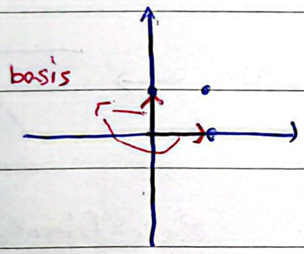
Determinant of a product: $\det(AB) = \det(A)\det(B)$

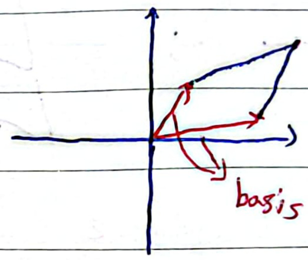
☆ product of a singular and a non-singular matrix is singular

Determinant of an inverse: $\det(A^{-1}) = \dfrac{1}{\det(A)}$

Bases (Plural of basis):

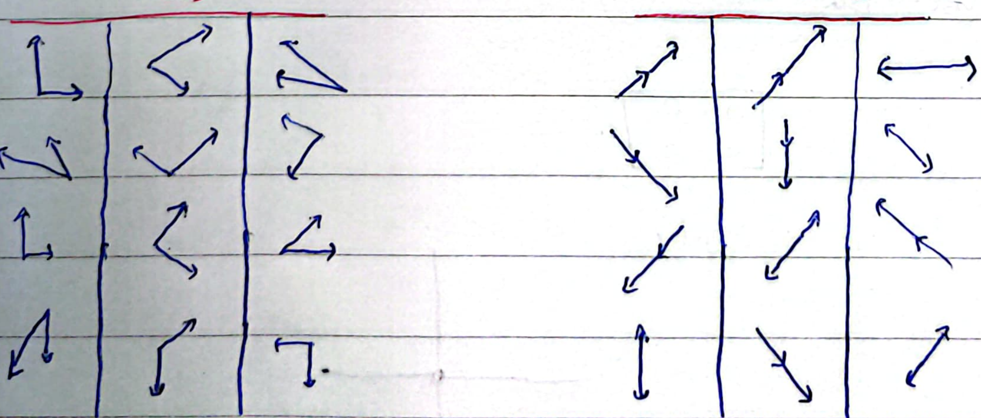
$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$$





☆ every point in the space can be expressed as a linear combination of elements in the basis

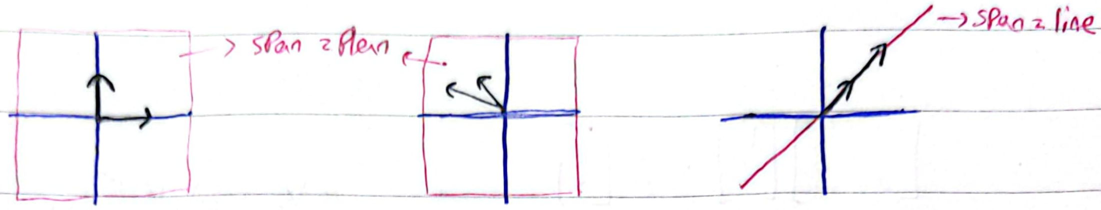
bases | Not bases

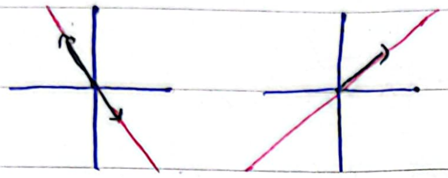


☆ We can go forward or backward with the same angle of the vector in basis
→ we can't change angle (only 0° or 180°)

Date: Subject:

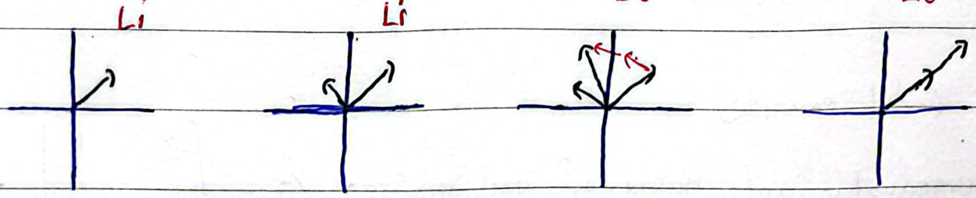
Span: The span of a set of vectors is the set of points that can be reached by walking in the direction of these vectors in any combination





★ a basis is a minimal spanning set

Linearly independent (Li) and linearly dependent (Ld) vectors: a group of vectors is said to be linearly independent if none of the vectors in the group can be obtained as linear combination of the others

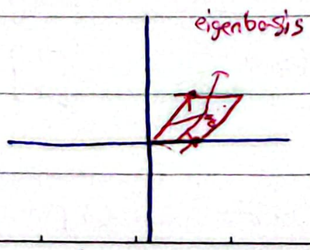


★ if you have more vectors than dimension of the space your trying to span, → Ld

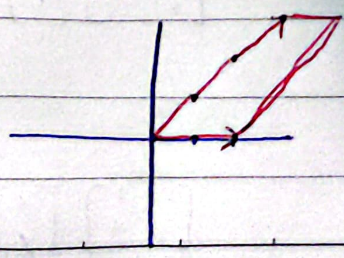
Basis:
- 1) spans a vector space
- 2) vectors are linearly independent

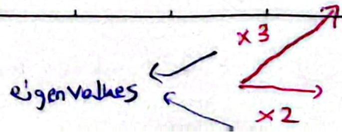
Eigenbasis:

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



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





→ a basis that sends a parallelogram to another parallelogram with sides parallel to the original one.

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

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

eigen Values

eigen Vectors

$$\underbrace{A v_1}_{\text{Matrix Mult (More Work)}} = \underbrace{\boxed{\lambda_1 v_1}}_{\text{scalar Mult (less work)}} \quad \text{Pair1}$$

$$A v_2 = \boxed{\lambda_2 v_2} \quad \text{Pair2}$$

★ we can use matrix eigenvectors to turn a large computation into a smaller one

★ ★ ★

eigenvectors: direction of stretch

eigenvalues: How much stretch

Finding eigenvalues (with example)

if λ was an eigenvalue and Matrix of Transformation, $A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$:

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 2-\lambda & 1 \\ 0 & 3-\lambda \end{bmatrix} \rightarrow \text{should have } \det = 0$$

$$\rightarrow (2-\lambda)(3-\lambda) - 0 = 0 \rightarrow \begin{cases} \lambda_1 = 2 \\ \lambda_2 = 3 \end{cases}$$

Finding eigenvectors (with example above) $\lambda_1 = 2$ $\lambda_2 = 3$

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2 \begin{bmatrix} x \\ y \end{bmatrix} \rightarrow 2x + y = 2x \rightarrow x = 1 \text{ (can be)}$$

$$3y = 2y \rightarrow y = 0 \qquad \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

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

$$3y = 3y \rightarrow y = 1 \text{ (can be)}$$

★ eigenvalues/vectors are only for square matrices.

★ for 2x2 : $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ for 3x3: $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$

if $\lambda_1 \neq \lambda_2 \rightarrow$ 2 eigenvectors (different direction)

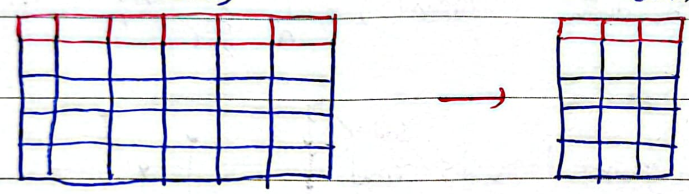
if $\lambda_1 = \lambda_2 \rightarrow$
- 1 eigenvector (1 direction)
- 2 " " (2 " ")

if $\lambda_1 \neq \lambda_2 \neq \lambda_3 \rightarrow$ 3 eigenvectors (3 dir) different

if $\lambda_1 = \lambda_2 \neq \lambda_3$
- 3 ev (3 dir)
- 2 ev (2 dir)

if $\lambda_1 = \lambda_2 = \lambda_3$
- 3 ev (3 diff dir)
- 2 ev (2 " ")
- 1 ev (1 dir)

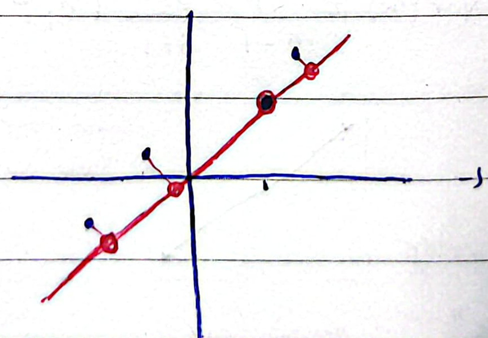
Dimensionality reduction : Reduce dims (# of cols) of dataset while preserving as much as info as possible



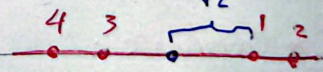
why? 1- leads to smaller dataset 2- Easier to visualize

Projection

x	y
1	1
1.2	1.6
-0.5	0.2
-1.3	-0.6



$\rightarrow$



Final cordinates on the line

$$\begin{bmatrix} 1 & 1 \\ 1.2 & 1.6 \\ -0.5 & 0.2 \\ -1.3 & -0.6 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1}{\sqrt{2}} = \begin{bmatrix} \frac{1+1}{\sqrt{2}} \\ \frac{1.2+1.6}{\sqrt{2}} \\ \frac{-0.5+0.2}{\sqrt{2}} \\ \frac{-1.3-0.6}{\sqrt{2}} \end{bmatrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix}$$

2d $\rightarrow$ 1d

SAMAND

★ To project a mat A onto a vector V

$$\underset{r\times 1}{A_p} = \underset{r\times c}{A}\ \underset{c\times 1}{\frac{V}{\|V\|_2}}$$

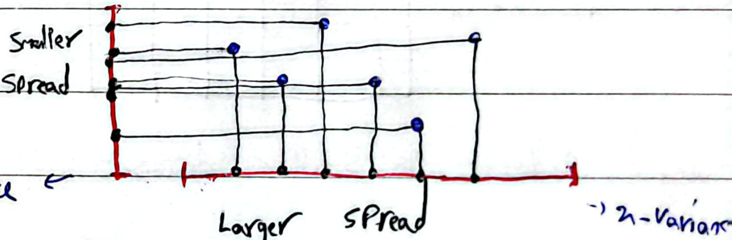
★ To project a mat A onto vectors V_1, V_2

$$\underset{r\times 2}{A_p} = \underset{r\times c}{A} \cdot \begin{bmatrix} \frac{V_1}{\|V_1\|_2} & \frac{V_2}{\|V_2\|_2} \end{bmatrix}$$

PCA (Principal component Analysis): Find the projection that preserves the max possible spread in your data

★ Mean: The average of the data $\quad \text{Mean}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{Mean}(y) = \frac{1}{n}\sum_{i=1}^{n} y_i$

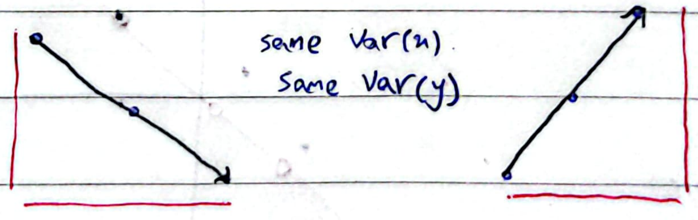
★ Variance: How spread out your data is



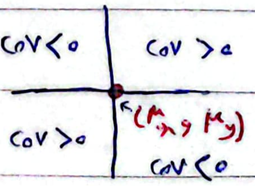
$$\text{Var}(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overset{\text{Mean}}{\mu_x})^2$$

★ Covariance:

" The dir of the relationship between two variables"



Negative cov $\qquad$ Positive cov



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

Covariance Matrix

$$C = \begin{array}{c|cc} & x & y \\ \hline x & Var(x) & cov(x,y) \\ y & cov(y,x) & Var(y) \end{array}$$

☆ $cov(x,x) = Var(x)$

☆ $cov(x,y) = cov(y,x)$

$$= \begin{bmatrix} cov(x,x) & cov(x,y) \\ cov(y,x) & cov(y,y) \end{bmatrix}$$

Formula to get C:

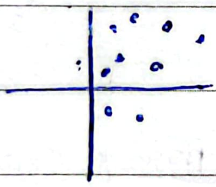
1- Arrange data with a dif feature in each column $A = \begin{bmatrix} x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{bmatrix}$

2- Calculate column averages

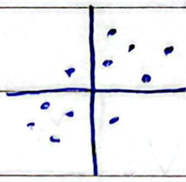
3- Subtract each avg from their respective col to generate $A - \mu$

4- $C = \frac{1}{n-1} (A-\mu)^T (A-\mu)$

PCA :



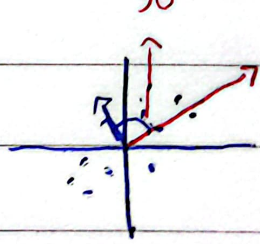
$\xrightarrow[\text{it's mean Point}]{\text{centered around}}$



$\xrightarrow{\text{find cov mat}} C = \begin{bmatrix} 9 & 4 \\ 4 & 3 \end{bmatrix}$

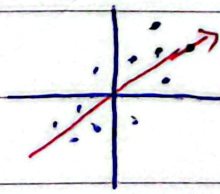
$\xrightarrow[\text{of cov mat}]{\text{find eigenvalues / vectors}}$

vector	$\begin{bmatrix} 2 \\ 1 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 2 \end{bmatrix}$
value	11	1



☆ eigenvectors of every mat that is symmetric around it's diagonal would have 90° orthogonal

$\xrightarrow[\text{the larger value}]{\text{choose eigenvector with}}$ $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ 11



$\xrightarrow[\text{on line}]{\text{Project data}}$

(2d → 1d).

☆ n-d ⟶ m-d = choose first m eigen~~values~~ vectors with biggest eigenvalues

PCA Formulation

imagine : You have n observation of 5 variables $(x_1, \dots, x_5)$ Goal : 5d → 2d

1) create matrix

$$X = \begin{bmatrix} x_{11} & \cdots & x_{15} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{n5} \end{bmatrix}$$

2) Center the data

$$X - \mu = \begin{bmatrix} x_{11} - \mu_1 & \cdots & x_{15} - \mu_5 \\ \vdots & & \vdots \\ x_{n1} - \mu_1 & & x_{n5} - \mu_5 \end{bmatrix}$$

3) Calc cov mat

$$C = \frac{1}{n-1} (X - \mu)^T (X - \mu)$$

4) Cal eigen values/vectors

λ_1 V_1 large
λ_2 V_2
λ_3 V_3
λ_4 V_4
λ_5 V_5 small

5) create projection matrix

$$V = \begin{bmatrix} \dfrac{V_1}{\|V_1\|_2} & \dfrac{V_2}{\|V_2\|_2} \end{bmatrix}$$

6) Project centered data

$$X_{PCA} = (X - \mu) V$$