Linear Algebra

(Maths for ML)

Input

$$y = m x + b$$

Output ← y; m → weight (w); x → Input; b → bias

$$y^{(1)} = w_1 x_1^{(1)} + w_2 x_2^{(1)} + \cdots + w_n x_n^{(1)} + b$$

$$y^{(2)} = w_1 x_1^{(2)} + w_2 x_2^{(2)} + \cdots + w_n x_n^{(2)} + b$$

$$\vdots$$

no. of records ← m

$$y^{(m)} = w_1 x_1^{(m)} + w_2 x_2^{(m)} + \cdots + w_n x_n^{(m)} + b$$

$\Rightarrow$ x and y are different for each equation but

$w_1, w_2, \ldots, w_n, b$ are same for each equation.

$$\begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix} \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \cdots \\ x_1^{(2)} & x_2^{(2)} & \cdots & \\ \vdots & & & \\ x_1^{(m)} & \cdots & & \end{bmatrix} = \begin{bmatrix} y^{(1)} & y^{(2)} & \cdots & y^{(m)} \end{bmatrix}$$

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

System of Linear Equations

System of Sentences :

System 1	System 2	System 3
The dog is black The cat is orange.	The dog is black The dog is black.	The dog is black The dog is white
Complete	Redundant	Contradictory

$\underbrace{\text{System 1}}_{\text{Non singular (most informative)}}$ $\underbrace{\text{System 2, System 3}}_{\text{singular (less informative)}}$

(how much redundant → rank)

ex.

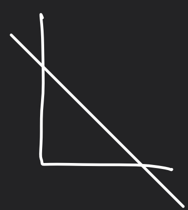
① $a + b = 10$

$a + 2b = 12$ (Unique sol^n)

so $b = 2$ and $a = 8$

② $a + b = 10$

$2a + 2b = 20$



not enough info

(∞ solutions)

③ $a + b = 10$

$2a + 2b = 24$

(No solution)

Non - linear equations:

$$a^2 + b^2 = 10$$

$$\sin(a) + b^5 = 15$$

$$2^a - 3^b = 0$$

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

Line : — 2 variable

Plane : — 3 variable

=> Turn constant to 0 and then check singular or Non - singular so constant have no role in S & NS determination.

System of equation as matrices :

$$a + b = 0$$

$$a + 2b = 0$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} [a \quad b] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

(NS)

Linear Dependence between rows:

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

$R3 = R1 + R2$

Linearly dependent (Singular)

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

$R2 = \frac{R3 + R1}{2}$

L.D (Singular)

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

Linearly Independent

Non singular

Determinant:

If $D = 0 \Rightarrow$ singular

If $D \neq 0 \Rightarrow$ Non-singular

Matrix Row Reduction :

$$\begin{vmatrix} 5 & 1 \\ 4 & -3 \end{vmatrix} \dashrightarrow \begin{vmatrix} 1 & 0.2 \\ 0 & 1 \end{vmatrix} \dashrightarrow \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$$

Upper diagonal matrix (Row Echelon form)

Reduced Row Echelon form

Row Echelon form :

$$\begin{vmatrix} 1 & * & * & * & * & * \\ 0 & 1 & * & * & * & * \\ 0 & 0 & 1 & * & * & * \\ 0 & 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 0 & \ddots & * \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

$\Rightarrow$ Row operations preserve singularity.

Interchanging rows $\longrightarrow$ - det

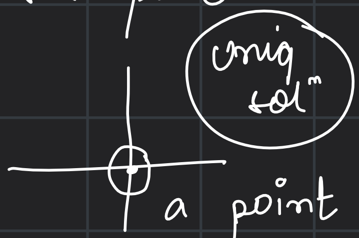
if one row addition $\longrightarrow$ det

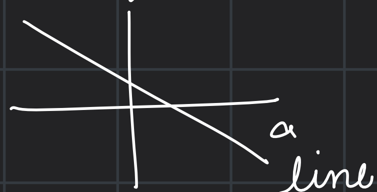
if multiplied one row with scalar $\rightarrow$ Scalar $\times$ det.

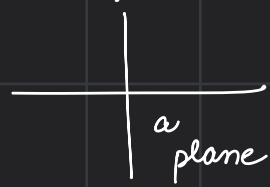
Rank of a matrix:

- how much info a system carries
- compressing images.

Sys 1	Sys 2	Sys 3
$a + b = 0$	$a + b = 0$	$0a + 0b = 0$
$a + 2b = 0$	$2a + 2b = 0$	$0a + 0b = 0$
Rank = 2	Rank = 1	Rank = 0
Dim^n of sol space = 0	Dim of sol space = 1	Dim of sol space = 2







Rank = No. of rows − Dimension of solution space

if Rank < no. of rows :- Non singular

Row Echelon Form :

$$\begin{vmatrix} 5 & 1 \\ 4 & -3 \end{vmatrix} \longrightarrow \begin{vmatrix} 1 & \overset{0.2}{1/5} \\ 1 & \underset{-0.75}{-3/4} \end{vmatrix} \longrightarrow \begin{vmatrix} 1 & 0.2 \\ 0 & -0.75 \end{vmatrix}$$

$$\downarrow$$

Row Echelon form $\begin{vmatrix} 1 & 0.2 \\ 0 & 1 \end{vmatrix}$

(NS) Rank = 2

$$\begin{vmatrix} 5 & 1 \\ 10 & 2 \end{vmatrix} \longrightarrow \begin{vmatrix} 1 & 0.2 \\ 1 & 0.2 \end{vmatrix} \longrightarrow \begin{vmatrix} 1 & 0.2 \\ 0 & 0 \end{vmatrix}$$

RE form

Rank = 1
singular

Rules of Row Echelon form:

→ zeroes at bottom

→ each row has a pivot (left-most non-zero entry)

→ every point is to the right of pivots on the row above.

→ Rank of matrix is the number of pivots.

$$\begin{vmatrix} 2 & * & * & * & * & * \\ 0 & 1 & * & * & * & * \\ 0 & 0 & 3 & * & * & * \\ 0 & 0 & 0 & -3 & * & * \\ 0 & 0 & 0 & 0 & \ddots & * \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

pivots

Reduced Row echelon form:

$$\begin{vmatrix} 5 & 1 \\ 4 & -3 \end{vmatrix} \longrightarrow \underbrace{\begin{vmatrix} 1 & 0.2 \\ 0 & 1 \end{vmatrix}}_{\text{Row Echelon form}} \xrightarrow{R1 = R1 - R2 \times 0.2} \underbrace{\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}}_{\text{Reduced Row Echelon form}}$$

Rules of Reduced Row Echelon form:

→ is in Row Echelon form

→ each pivot is a 1.

→ any number above pivot is 0.

→ Rank = number of pivots.

Reduced Row Echelon form:

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

Gaussian Elimination Algo :

augmented matrix $\left[\quad\quad\right]\left[\quad\right] \Rightarrow \left[\begin{array}{c|c} \quad\quad & \quad \end{array}\right]$

$$\begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} \left[\begin{array}{ccc|c} 2 & -1 & 1 & 1 \\ 2 & 2 & 4 & -2 \\ 4 & 1 & 0 & -1 \end{array}\right]$$

1. Take R1 as pivot and make it 1.
2. Reduce the below pivots to zero

[Perform for each row]

Now R_3, we have pivot

Now start converting above pivot to 0.

Then do Back substitution

if
$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 5 \\ 0 & 0 & -1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right]$$
then infinite (∞) solⁿ.

if
$$\left[\begin{array}{ccc|c} - & - & - & - \\ - & - & - & - \\ 0 & 0 & 0 & 4 \end{array}\right]$$
then no solution.

⇒ Vector Algebra

Row Vector

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

$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ Column Vector

L1 norm : $||x||_1 = |x_1| + |x_2| + \dots + |x_n|$

L2 norm : $||x||_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} = \sqrt{(\text{dot pdt})}$

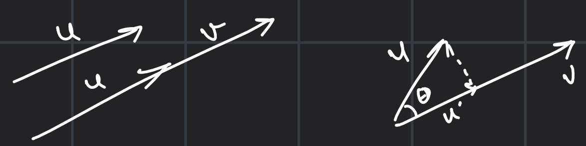
Diff / sum of vectors
└ component by component

Multiply by scalar
└ element wise

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

$\Rightarrow$ Orthogonal vectors have dot product $= 0$.

Dot product :



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

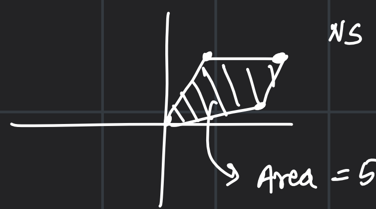
(diff in vectors & matrix?)

Matrix Inverse :

$$A \cdot A^{-1} = I$$

if det $= 0$ $\Rightarrow$ no inverse

$$\begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix} \qquad \text{Det} = 5$$



if Det $= 0$

singular

if $D < 0$ then counter-clockwise order

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

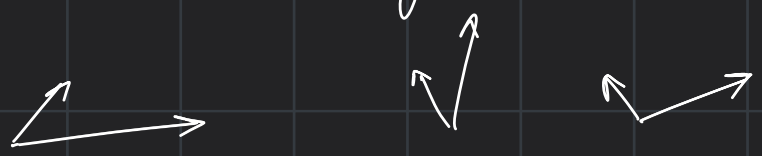
when A is invertible:

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

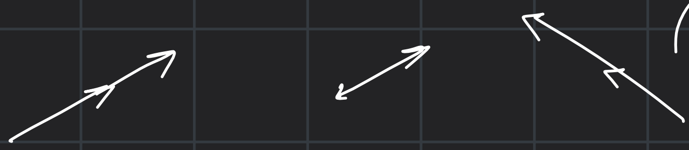
Bases (plural of "basis")

↳ any pair of vectors which can be used to reach to any point from the bases.

eg.

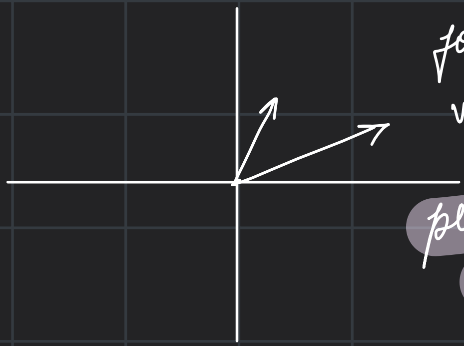


Not bases :

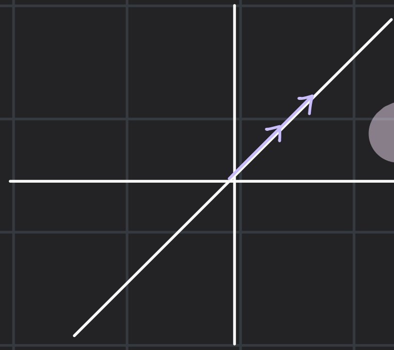


(in same direction)

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.

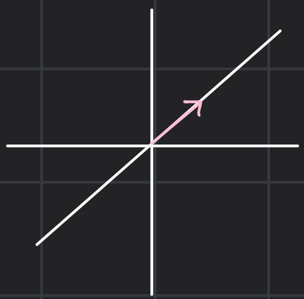


for these 2 vectors, whole plane is span



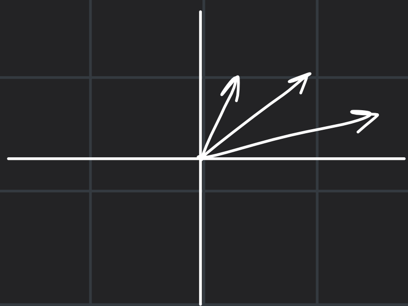
for these, line is span

Basis is a minimal spanning set.



this is basis for line

but this not



these are 3, one of them is redudant so not basis.

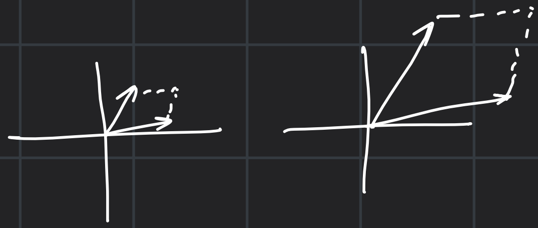
Number of element in basis is dimension.

Linear dependence: $\alpha v_1 + \beta v_2 + \gamma v_3 = 0$

<u>Basis</u> : a set of vectors that:

- spans a vector space
- is linearly dependent

Eigenvalues / vectors:



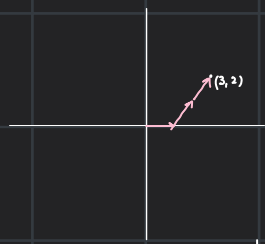
in same direction, how much stretch

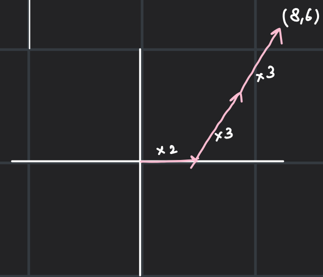
$$A_1 v_1 = \underline{\lambda_1} V_1$$

scalar multiplication
(less computation)

eg

$$\begin{vmatrix} 2 & 1 \\ 0 & 3 \end{vmatrix} \begin{vmatrix} 3 \\ 2 \end{vmatrix} = \begin{vmatrix} 8 \\ 6 \end{vmatrix}$$





direction of stretch ← e. vector → $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

how much stretch ← e. values → 2, 3

Eigen basis → set of matrix's eigenvector, can be arranged as a matrix with one eigen vector in each column.

$$\boxed{\det(A - \lambda I) = 0}$$ Characteristic Equation

eg.

$$\begin{vmatrix} 9 & 4 \\ 4 & 3 \end{vmatrix} \qquad \det \begin{vmatrix} 9-\lambda & 4 \\ 4 & 3-\lambda \end{vmatrix} = 0$$

$$(9-\lambda)(3-\lambda) - 16 = 0$$

$$\boxed{\lambda = 11, 1}$$

A

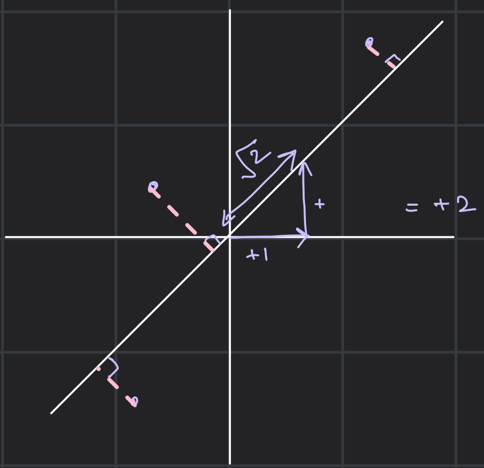
if repeated eigenvalues
↓
either different direction or same.

$$(A - \lambda I)x = 0$$

$x \to$ eigenvector.

Dimensionality Reduction :

Role of PCA : reduce # of columns of dataset while preserving as much info possible.
(Principal Component Analysis)



$$\sqrt{2} = \frac{1+1}{\sqrt{2}} \rightarrow$$ overshoot by $\sqrt{2}$ so divide by $\sqrt{2}$ (L_2 norm)

so $$\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1+1}{\sqrt{2}}$$

or $$\frac{1}{\left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\|_2}$$

To project a matrix A onto a vector v:

$$A_p = A \overbrace{\frac{v}{\|v\|_2}}^{V} \quad \text{or} \quad A \overbrace{\begin{bmatrix} \frac{v_1}{\|v_1\|_2} & \frac{v_2}{\|v_2\|_2} \end{bmatrix}}^{V}$$

$$A_p = AV$$

PCA : more spread, more info.
: reduces dimensions while minimizing info loss.
: simpler viz.

Mean :

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

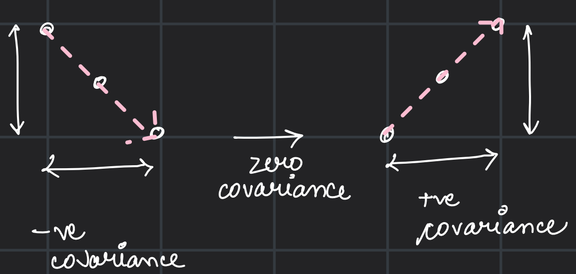
$$\text{Mean}(y) = \frac{1}{n} \sum_{i=1}^{n} y_i$$

Variance : $\propto$ spread

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

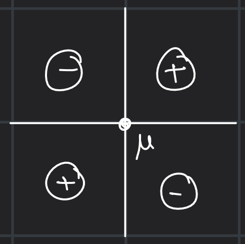
" The average square distance from the mean "

Problem:



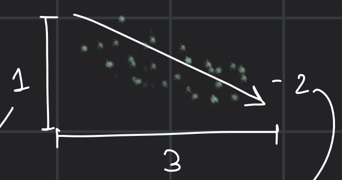
different distance but same spread.

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

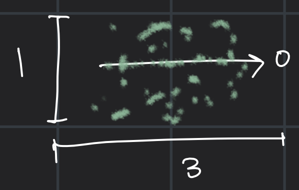


so covariance is checking if there are more points in − or + quadrants.

eg.

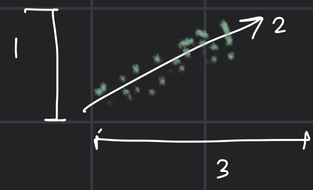


$$\begin{bmatrix} 3 & -2 \\ -2 & 1 \end{bmatrix}$$



$$\Downarrow$$

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



$$\Downarrow$$

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

$$C = \begin{array}{c c} & \begin{array}{cc} x & y \end{array} \\ \begin{array}{c} x \\ y \end{array} & \begin{vmatrix} \text{Var}(x) & \text{Cov}(x,y) \\ \text{cov}(x,y) & \text{Var}(y) \end{vmatrix} \end{array}$$

$\text{Cov}(x, y) = \text{Cov}(y, x$

$\text{Cov}(x, x) = \text{Var}(x)$

$$A = \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{vmatrix} \qquad \mu = \begin{vmatrix} \mu_x & \mu_y \\ \vdots & \vdots \\ \mu_x & \mu_y \end{vmatrix}$$

$$C = \frac{1}{n-1} \underset{2 \times n}{(A - \mu)^T} \underset{n \times 2}{(A - \mu)}$$

Best line = most covariance.

The eigenvector with largest eigenvalue will have more spread in that direction.

why PCA? - any point and its projection will stretch between the value of biggest and smallest eigenvalue.

(1) Create matrix

Goal: Reduce to 2 variables

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{15} \\ & & & \end{bmatrix}$$

(2) Center the data

$$X - \mu = \begin{vmatrix} x_{11}-\mu_1 & x_{12}-\mu_2 & \cdots & x_{15}-\mu_5 \\ x_{21}-\mu_1 & x_{22}-\mu_2 & \cdots & \\ \vdots & \vdots & & \\ x_{n1}-\mu_1 & x_{n2}-\mu_2 & \cdots & \end{vmatrix}$$

(3) Calculate covariance matrix:

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

(4) Calculate eigenvalue and eigenvectors

Big ↑	λ_1	v_1
	λ_2	v_2
	$\vdots$	
small	λ_5	v_5

(5) Create projection matrix

$$V = \begin{vmatrix} \frac{v_1}{\|v_1\|_2} & \frac{v_2}{\|v_2\|_2} \end{vmatrix}$$

(6) Project centered data

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

Discrete Dynamical Systems :

If today is

prob of next cloudy day

$$\begin{array}{c|ccc} & \text{Sunny} & \text{Cloudy} & \text{Rainy} \\ \hline \text{Sunny} & 0.80 & 0.45 & 0.30 \\ \text{Cloudy} & 0.15 & 0.35 & 0.40 \\ \text{Rainy} & 0.05 & 0.20 & 0.30 \end{array} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.35 \\ 0.20 \end{bmatrix}$$

→ next day prob.

(+) = 1

$$\begin{array}{c|ccc} & \text{Sunny} & \text{Cloudy} & \text{Rainy} \\ \hline \text{Sunny} & 0.80 & 0.45 & 0.30 \\ \text{Cloudy} & 0.15 & 0.35 & 0.40 \\ \text{Rainy} & 0.05 & 0.20 & 0.30 \end{array} \begin{bmatrix} 0.45 \\ 0.35 \\ 0.20 \end{bmatrix} = \begin{bmatrix} * \\ * \\ * \end{bmatrix}$$

→ next to next day prob.

(+) = 1

Transition / Markov Matrix

if we keep on doing this.

we will get a vector which won't update further.

with eigenvalue = 1 called **Equilibrium Vector**