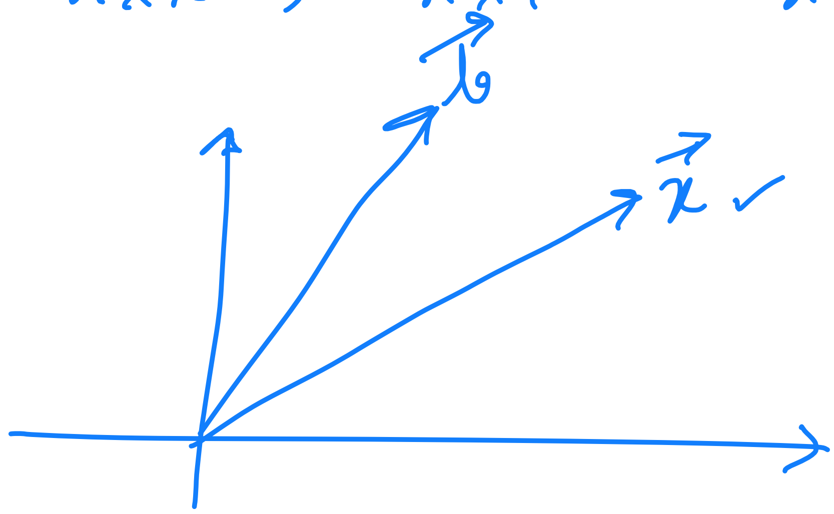
Eigenvalues & Eigenvectors

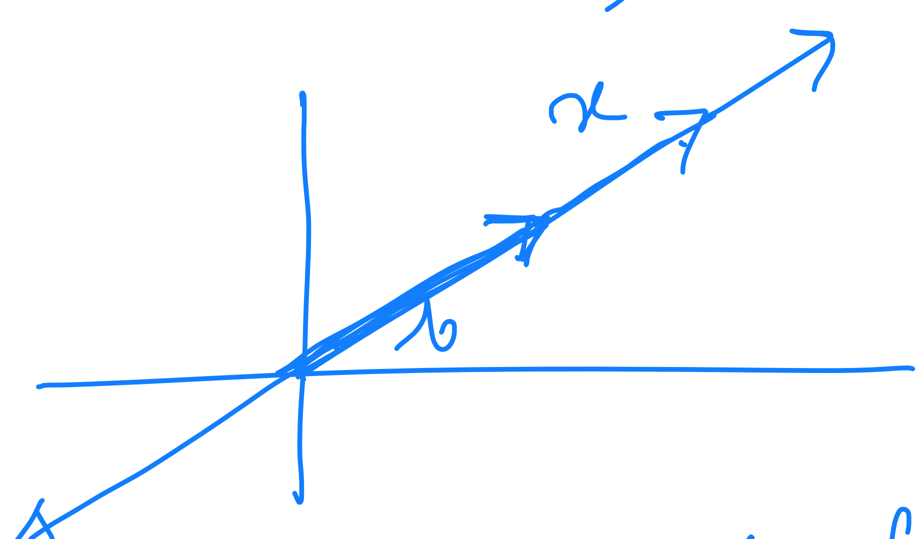
— Linear equation of the form $\boxed{A\vec{x} = \vec{b}}$

$A_{n \times n}$, $x_{n \times 1}$, $b_{n \times 1}$



$\vec{x} \longrightarrow \boxed{A_{n \times n}} \longrightarrow A\vec{x} = \vec{b}$

— If there are $\vec{x}$ vector for matrix A that does not change the direction, then what happens?



$\vec{x} \longrightarrow \boxed{A} \longrightarrow \vec{b} = A\vec{x}$

The mathematical formulation for this condition when $\vec{b}$'s orientation/direction is same as $\vec{x}$. is represented as

$$A\vec{x} = \lambda \vec{x}$$

λ : constant that represent amount of stretch / shrinkage the attributes go through in $\vec{x}$ direction.

$\vec{x}$ is called the eigenvector

& λ is called the eigenvalue.

Find eigenvalues λ & Eigenvectors $\vec{x}$.

A : $n \times n$ matrix

I : $n \times n$ matrix (Identity matrix)

$$Ax = \lambda x$$

$$Ax - \lambda . I x = 0$$

$$\boxed{(A - \lambda I) x = 0} \longrightarrow$$

Null space

$A' B = 0$ ✓

To find $\vec{x} \neq 0$

$\Rightarrow x$ is in the null space of $(A - \lambda I)$

- from rank-nullity theorem.

$$\underset{(A-\lambda I)}{\text{rank}} + \text{nullity}(A - \lambda I) = n$$

$\Rightarrow$ rank$(A-\lambda I)$ has to be $< n$

$\Rightarrow$ nullity has to be atleast 1.

$\Rightarrow$ ensures non-trivial solution

$\Rightarrow$ $|A - \lambda I| = 0$ solve the system of linear equations to obtain eigenvalues λ and corresponding eigenvector

Ex: Find eigenvalues & eigenvectors for

$$A = \begin{bmatrix} 8 & 7 \\ 2 & 3 \end{bmatrix}$$

$$|A - \lambda I| = 0$$

$$\left| \begin{bmatrix} 8 & 7 \\ 2 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

Or, $\begin{vmatrix} 8-\lambda & 7 \\ 2 & 3-\lambda \end{vmatrix} = 0$

Or, $(8-\lambda)(3-\lambda) - 14 = 0$

Or, $24 - 3\lambda - 8\lambda + \lambda^2 - 14 = 0$

Or, $\lambda^2 - 11\lambda + 10 = 0$

Or, $(\lambda - 10)(\lambda - 1) = 0 \Rightarrow$ eigenvalues $\boxed{\lambda = 10, 1}$

Find eigenvectors: $\quad Ax = \lambda x$

$\lambda = 1,$ $\begin{bmatrix} 8 & 7 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 1 \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$8x_1 + 7x_2 = x_1$

$2x_1 + 3x_2 = x_2$

Or, $7x_1 + 7x_2 = 0 \Rightarrow x_1 + x_2 = 0$

$2x_1 + 2x_2 = 0 \Rightarrow x_1 + x_2 = 0$

$x_1 + x_2 = 0 \Rightarrow x_1 = -x_2$

represent eigenvector as unit vector

$\vec{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} k \\ -k \end{bmatrix}$ [square root of sum squares of components is equal to 1 = unit vector]

$\sqrt{k^2 + (-k)^2} = \sqrt{1}$

$\sqrt{2k^2} = 1$

$k = \frac{1}{\sqrt{2}}$

for $\lambda=1, \vec{x} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -1/\sqrt{2} \end{bmatrix}$

for $\lambda = 10$; $\begin{bmatrix} 8 & 7 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 10 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$8x_1 + 7x_2 = 10x_1$$

$$2x_1 + 3x_2 = 10x_2$$

Or, $\cancel{-}2x_1 = \cancel{-}7x_2$

$$2x_1 = 7x_2$$

Or, $x_2 = \frac{2}{7}x_1$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} k \\ \frac{2k}{7} \end{bmatrix}$$

unit vector,

$$\sqrt{k^2 + \left(\frac{2k}{7}\right)^2} = 1$$

$$\sqrt{k^2 + \frac{4k^2}{49}} = 1$$

$$\sqrt{\frac{53k^2}{49}} = 1$$

$$\frac{k}{7}\sqrt{53} = 1$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{7}{\sqrt{53}} \\ \frac{2}{\sqrt{53}} \end{bmatrix} \text{ for } \underline{\lambda = 10}.$$

Properties:

→ If $A = A^T$ (Symmetric matrix) eigenvalues are real values

→ eigenvectors corresponding to distinct λ values are orthogonal

⇒ eigenvectors are linearly independent

for $A = A^T$ eigenvectors are Linearly Independent

- Let A be a symmetric matrix assume r eigen values are 0 out of n

$\Rightarrow$ dimension of null space is r.
from rank-nullity theorem,
rank is $n-r$.

$\Rightarrow$ using these $(n-1)$ linearly independent vectors row space of the matrix can be obtained.
$\Rightarrow$ basis vectors for the column space.

X — X

Dimensionality Reduction:

1) PCA
Principal component Analysis

2) Singular value Decomposition (SVD)

3) LDA
Linear Discriminant Analysis

$A_{m \times n} \rightarrow$ lower dimension $p < n$
so that loss of information is minimized

$\boxed{\text{PCA}}$: Principal Component Analysis.

→ is a statistical tool that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components

→ linearly transform the data into a new coordinate system, where most of the variances of the data can be described with fewer dimension

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

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

Covariance : measure the direction of relationship between two variables.

→ +ve covariance both variables tend to be high / low at the same time

→ -ve , one variable high, the other low or vice versa.

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

$\bar{x}, \bar{y}$ means.

Principal components of a collection of points are a sequence of unit vectors, where the i^{th} vector is the direction of a line that best fits the data while being

Orthogonal to the first $i-1$ vectors.

- Principal components are orthogonal basis set. $\Rightarrow$ new coordinate system.
- First principal component has the largest possible variance and each succeeding component has in turn the highest variance possible under the constraint that it is orthogonal to preceding components.

	x_1	x_2	x_3
x_1	var		
x_2		var	
x_3			var

How it is done:

PCA

✓ i) Eigenvalue decomposition of data covariance matrix (symmetric matrix) obtained after mean centering

ii) Singular value decomposition of data matrix, after mean centering & normalization

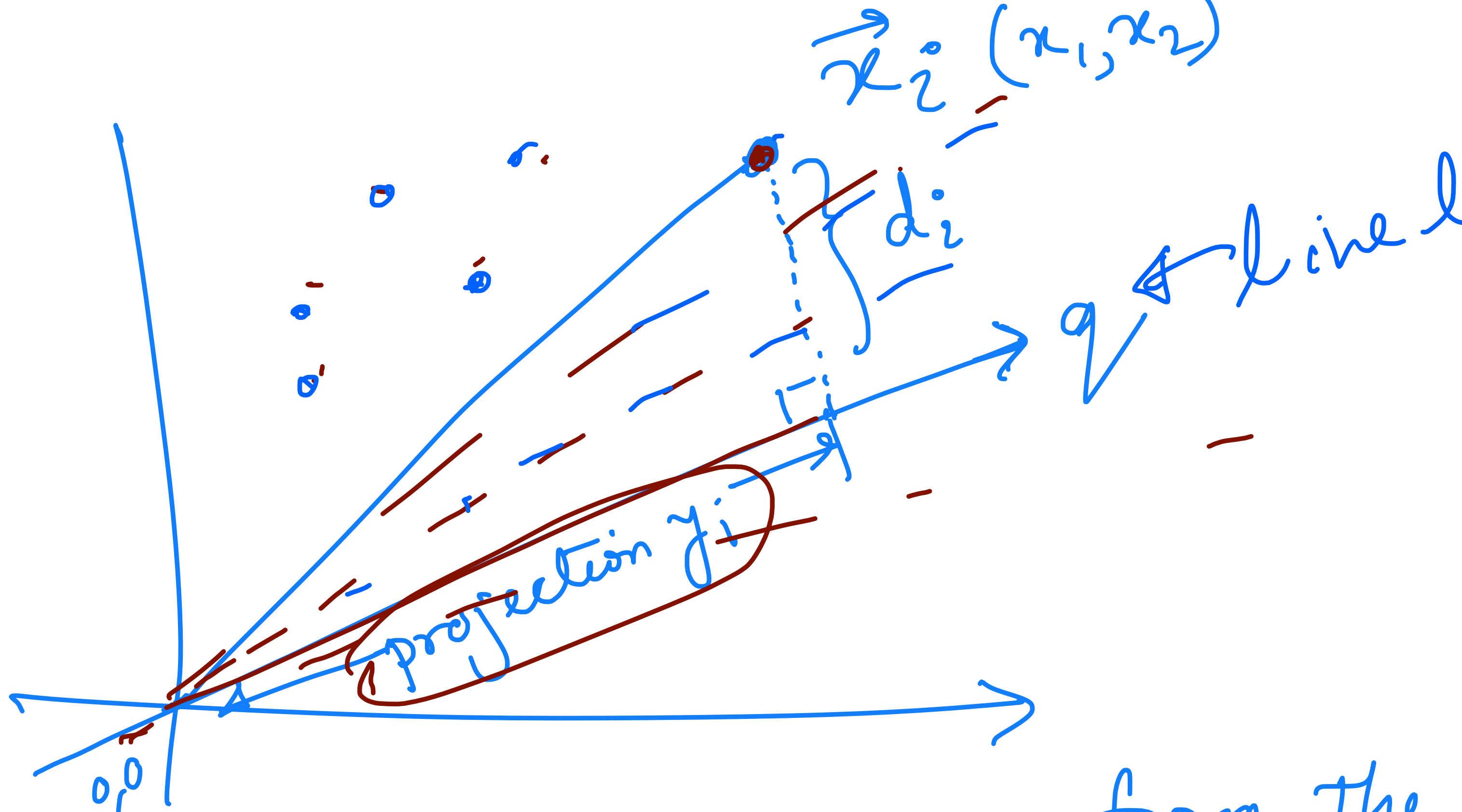
i) Maximization of Variance Estimate:

Let $X = \{ \vec{x}_1, \dots \vec{x}_i, \dots \vec{x}_n \}$ each $x_i \in \mathbb{R}^m$ be a set of points such that the centroid of X is the origin, ie the data is mean centered.

$$(x_{1j} - \vec{x}_1) \Rightarrow \text{mean centered.}$$

$$\begin{array}{c} \\ \vec{x}_1 \\ \vec{x}_2 \\ \vec{x}_3 \\ \vdots \\ n \end{array} \begin{array}{c} \qquad\qquad m \\ \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ & & & \\ & & & \\ & & & \end{bmatrix} \end{array} \begin{array}{l} \to \bar{x}_1 \\ \text{mean} \\ \to \bar{x}_2 \\ \text{centred} \\ \to \bar{x}_n \end{array} \qquad \begin{bmatrix} x_{11} - \bar{x} & x_{12} - \bar{x} & \cdots & x_m - \bar{x} \\ & & & \\ & & & \end{bmatrix}$$

Q: Find a line through origin that maximizes the projections y_i of the points $\vec{x}_i$ on l



d_i : the distance of x_i from the vector q.

- Let $\vec{q}$ denote the unit vector along line l

- The projection y_i of x_i on l

$$y_i = x_i^T q. \quad \text{—(1)}$$

- The mean squared projection, which is the variance, V over all points is

$$V = \frac{1}{n} \sum_{i=1}^{n} y_i^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i^T q)^2 \geq 0 \quad \text{—(2)}$$

We know $x_i^T q = q^T x_i$, So

$$V = \frac{1}{n} \sum_{i=1}^{n} (q^T x_i)(x_i^T q)$$

$$= q^T \underbrace{\left[\frac{1}{n} \sum_{i=1}^{n} x_i x_i^T \right]}\, q \quad — (3)$$

Let $C = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T$ — (4) $\Rightarrow$ it is the $\boxed{\text{covariance matrix of the data points.}}$

C is a symmetric matrix since $cov(x, y)$, $x \neq y$ is same as $cov(y, x)$.

Q: How to find $\vec{q}$ ($\vec{q}$ is a unit vector that maximizes V)

So maximize $\boxed{V = q^T C q}$ — (5) from (3) & (4)

subject to $\|q\| = 1$ [i.e. $q \cdot q^T = 1$]
magnitude

$\Downarrow$

$q \cdot q^T - 1 = 0$

Use Lagrangian multiplier method:

→ combines V and the constraint using multiplier λ.

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

q $\quad q^T$

$1 + 2$

New equation becomes:

$$\text{maximize } V' = q^T C q - \lambda \underbrace{(q^T q - 1)}_{}$$ — (6)

$$\text{Find } \frac{\partial V'}{\partial q} = 2 q^T C - 2\lambda q^T = 0$$ — (7)

Rearranging & using $C = C^T$,

we get $q^T C = \lambda q^T$ — (8)

use transpose on both sides,

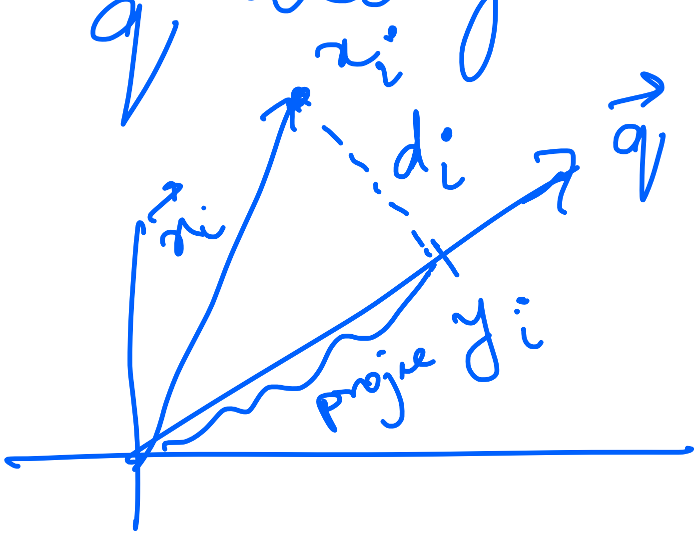
$$\boxed{C q = \lambda q}$$ ⇒ the eigenvector q of covariance matrix C gives the line that maximizes the variance V

replace (8) in (5)

$$V = q^T C q = q^T \lambda q = \lambda \quad (10)$$

$$\lambda > 0.$$

$\Rightarrow$ Finding eigenvalues & eigenvector corresponding to maximum eigenvalue gives the direction / unit vector $\vec{q}$ along which variance is maximum.



length of projection

unit vector

$$d_i + y_i \vec{q} = x_i$$

$$\vec{d_i} = \vec{x_i} - y_i \vec{q}.$$

$$\text{distance} = \|\vec{x}_i - y_i \vec{q}\| = \|d_i\| \quad (11)$$

we know $\underbrace{y_i = x_i^T q}$

$$\|d_i\| = \|x_i - \underbrace{y_i}\, q\| = \|\vec{x}_i - (x_i^T q) q\|$$

$$d_i^2 = (x_i - (x_i^T q).q)^T (x_i - (x_i^T q) q) \quad (12)$$

$$= x_i^T x_i - x_i^T (x_i^T q) q$$

$$- (x_i q^T) q^T x_i + (x_i q^T) q^T (x_i^T q) q$$

Since $x_i^T q$ is a scalar & $x_i^T q = q^T x_i$

$$d_i^2 = x_i^T x_i - (x_i^T q)^2$$

Arrange over all i points

$$D = \frac{1}{n} \sum_{i=1}^{n} d_i^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^T x_i - \boxed{13}$$

$\Rightarrow$ as we maximize variance

$\Rightarrow$ minimizes the mean squared distance D to the datapoints

x_2:

$$\left[\text{dist. of point to line } (d_i)\right]^2 + \left[\text{proj of } x_i \text{ on } q\ (y_i)\right]^2 = x_{i1}^2 + x_{i2}^2 + \ldots + x_{in}^2$$

$$d_i^2 = \boxed{(x_{i1}^2 + x_{i2}^2 + \cdots x_{in}^2)} \boxed{- \left(\begin{array}{c}\text{proj of } x_i \\ \text{on } q\ (y_i)\end{array}\right)^2} \quad (14)$$

Variance.

writing eq (13) properly.

$$D = \frac{1}{n}\sum d_i^2 = \underbrace{\frac{1}{n}\sum x_i^T x_i}_{} - V \quad (15)$$

$\Rightarrow$ since we have maximized variance V and V is subtracted in (15)

$\Rightarrow$ D is minimized

<u>Our objective in PCA:</u>

To find a vector q that maximizes the spread (V) also minimizes the mean distance of all the points to the vector q $\Rightarrow$ gives a good representation of the points in lower dimension