AFTER MIDSEM

EIGEN VALUE PROBLEMS IN OPTIMIZATION

eigen vectors are vectors x that do not get "rotated" by A, only "stretched" (by a factor of λ).

$$Ax = \lambda x$$

eigen value of A can be obtained by solving the characteristic equation $|A - \lambda I| = 0$. eigen vectors form the null space of matrix $(A - \lambda I)$

interesting properties

- $A^2 x = \lambda^2 x$

- $\prod_i \lambda_i = |A|$

- $\sum_i \lambda_i = tr(A)$

- every vector is an eigen vector of I.

- symmetric matrices have real eigen values.

- spectral theorm:
symmetric matrices $S = Q \Lambda Q^T$

- generalized eigen value problem of 2 symmetric matrices A and B.

$$Ax = \lambda Bx$$

$$\underset{n\times n}{S} = \sum_{i=1}^{N} \lambda_i X_i X_i^T$$

$$S = Q \lambda Q^T$$

$$\begin{bmatrix} & & \\ & Q & \\ & & \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \ \lambda_N \end{bmatrix} \begin{bmatrix} & & \\ & Q^T & \\ & & \end{bmatrix}$$

data compression:

new basis vectors, if you get eigen vectors, you can get more optimally.

donot want to transmit basis - DCT

$Ax = \lambda x$ classical eigen value problem.

$\downarrow$

$Ax = \lambda Bx$ generalized eigen value problem.

eigen - standard trick, do cholesky of B.

$Ax = \lambda L L^T x$

$L^{-1} A x = \lambda L^T x$

let $y = L^T x$

$x = L^{-T} y$

$(L^{-1} A L^{-T}) y = \lambda y$?? converted to classical problem.

$$Ax = \lambda L L^T x$$

$$L^{-1} A x = \lambda L^T x$$

$$A' y = \lambda y$$

$$A' = L^{-1} A L^{-T} \qquad y = L^T x$$

$$Ax = \lambda B x$$

$B^{-1}A$ need not have similar properties of B and A.

---

opt. $x^T A x$ s.t. $\|x\| = 1$

$$x^T A x - \lambda (x^T x - 1)$$

$\frac{\partial}{\partial x} = 0 \Rightarrow$

$Ax = \lambda x$ or x is EV of A.

as we know $x^T x = 1$

$$\Rightarrow x^T A x = \underline{\lambda}.$$

pick largest value, so objective is λ.
min., pick smallest eigen value.

$$Ax = 0$$

$$\min \|Ax\| \quad \text{st.} \quad \|x\| = 1$$

remove trivial solns.
get a fixed magnitude

scalar

$$x^T A^T A x - \lambda (x^T x - 1)$$

$$A^T A x = \lambda x$$

$$z^* = \lambda$$

PRESENCE OF EV IN OM

many optimization problems take the form:

$$\min_x z = x^T A x,$$

$$\text{st.} \quad \|x\|^2 = 1$$

soln. gives us $Ax = \lambda x$

$$z^* = \lambda$$

maximum variance line fitting

let $y_1 \dots y_N$ be the points,

and ω be the direction.

Variance ~~of~~ on x should be maximized with a constraint on $\|\omega\| = 1$. also $x = \omega^T(y - b)$

$$\frac{1}{N} \sum_{i=1}^{N} \left(x_i - \frac{1}{N} \sum_{i} x_i\right)^2 + \lambda\,(\omega^T\omega - 1)$$

least norm soln. to $Ax = 0$.

$$\min_{x} \|Ax\|^2$$

st. $\|x\|^2 = 1$

soln. $(A^TA)\,x = \lambda x$

$z^* = \lambda$.

APPLICATIONS: PCA

objective function of PCA is defined as:
given data matrix X: $(N \times M)$, find a direction

u such that the variance of the projection of X on u is maximized (max. information captured).

projection of X on u is X_u.

mean of projections is $\bar{X}_u$.

$$\max_u \|(X-\bar{X}).u\|^2, \quad \|u\|^2 = 1$$

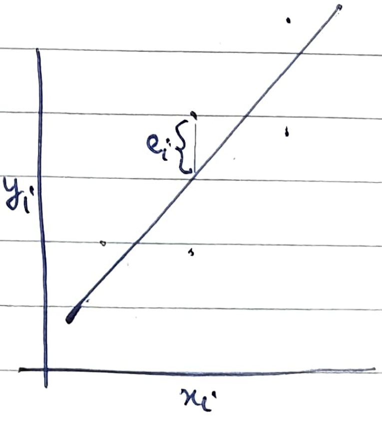
the constraint comes because u is only a direction

$$\max_u u^T \Sigma u, \quad u^T u = 1$$

$$\max_u u^T \Sigma u - \lambda (u^T u - 1)$$

$$\Sigma u = \lambda u, \quad z^* = \lambda$$

$$\min \|y_i - w^T x_i\|$$

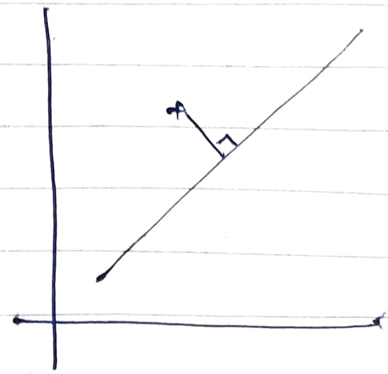


$L2 \longrightarrow$ LSE

$L1 \longrightarrow$ LP

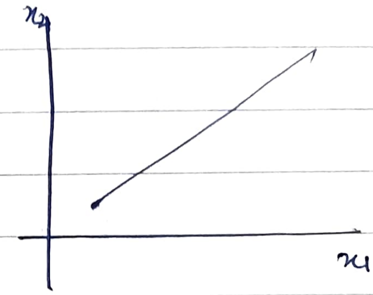
$L0 \longrightarrow$ most points passing

$L\infty \longrightarrow$ distance to largest outlier smallest.



suddenly becomes an EV problem.

(eg. PCA).



set of dimensions

$x_1 \ldots x_n$

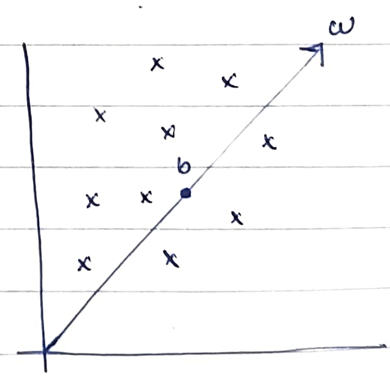
set of points

x_1, x_2, x_3

interested in finding the plane

no special axis

onthogonal distance.



Variance is maximum

$y_1 \ldots y_N$

$$x = \omega^T (y - b)$$

to find

we would like variance to be maximized.

$$\max_{\omega} \frac{1}{N} \sum_i \left(x_i - \frac{1}{N} \sum x_i\right)^2 \quad \text{st.} \quad \|\omega\| = 1$$

$$\max_{\omega} \frac{1}{N} \sum_i (x_i - \bar{x})^2 - \lambda (1 - \omega^T \omega)$$

$$\Rightarrow \frac{1}{N}\sum_i \left(\omega^T(y_i - b) - \frac{1}{N}\sum_i \omega^T(y_i - b)\right)^2$$

$$\Rightarrow \frac{1}{N}\sum_i \left(\omega^T\left(y_i - b - \frac{1}{N}\sum_i y_i \cancel{+} \frac{1}{N} N b\right)\right)^2$$

$$\Rightarrow \frac{1}{N}\sum_i \left(\omega^T\left(y_i - \frac{1}{N}\sum y_i\right)\right)^2$$

$$\Rightarrow \frac{1}{N}\sum_i \omega^T \underbrace{(y_i - \bar{y})(y_i - \bar{y})^T}_{\text{covariance of } y_i} \omega$$

$$\omega^T \Sigma \omega - \lambda(1 - \omega^T\omega)$$

ω is the $\overline{EV}$ of Σ.

ω is the eigen vector corresponding to largest eigen value of Σ.

b is the mean of y_i.

projection of points resulting max. variance. (b).

$f: (x-1)^2 + y^2 \qquad x + y = 2$

$L: (x-1)^2 + y^2 - \lambda(x - y - 2) = 0$

$\frac{\partial L}{\partial x} = 0 \Rightarrow 2(x-1) - \lambda = 0 \qquad 2(x-1) = \lambda$

$\frac{\partial L}{\partial y} = 0 \Rightarrow 2y - \lambda = 0 \qquad 2y = \lambda$

$(3/2, 1/2) \qquad \frac{\partial L}{\partial \lambda} = 0 \Rightarrow x + y = 2.$

SIMPLE GRAPH PARTITIONING / CLUSTERING

- matrices

  \# edges = m

  \# vertices = n

  - incidence matrix (J) : $m \times n$
  - degree matrix: diagonal (D) : $n \times n$
  - adjacence (A) : $n \times n \in \{0, 1\}$
  - laplacian (L) : $n \times n$ $L = D - A$;
    $L = J^T J$
  - affinity / weight (A / ω) : $n \times n$

- affinity matrix A;

  A_{ij} affinity of ith and jth node (eg. $A_{ij} = e^{-d(x_i, x_j)}$)

- ω: vector with ω_i being the "membership" of i vertex / sample into the cluster.

- optimization problem:

  max. $\omega^T A \omega$

  st. $\omega^T \omega = 1$

- ω is the eigen vector of A and characterize a good cluster.

SPECTRAL GRAPH PARTITIONING / CLUSTERING

- A be the adjacency matrix.

eg. (i) fully connected with all nodes have degree d;
(ii) 2 separate components of each have degree d_1.

- $Ax = y$; y_i is the sum of neighbours of node i.

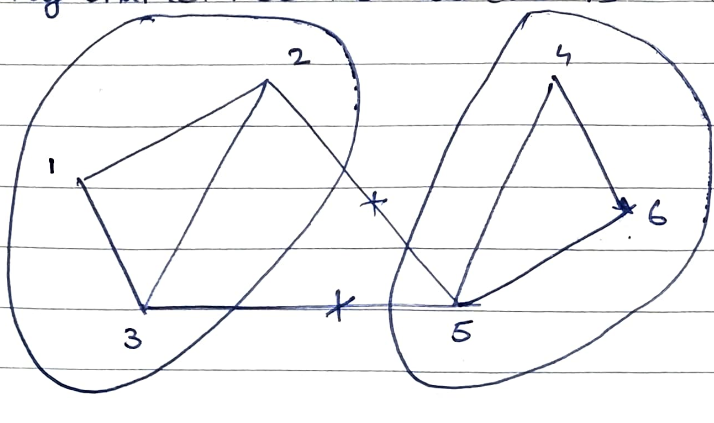
- case 1 : $Ax = \lambda x$; $x = [1, 1, \dots 1]^T$ and $\lambda = d$.

- case 2 : $x' = [1, \dots 1, 0, \dots 0]^T$ and $\lambda' = d_1$ and
$x'' = [0, \dots 0, 1, \dots 1]^T$ and $\lambda'' = d_1$

- $Lx = 0$ is true for a constant vector say $x_i = 1$ vector, with $\lambda = 0$ (why? D_{ii} is y_i).

- fiedler vector: eigen vector corresponding to the smallest positive (second smallest) eigen value.

- x_2 is orthogonal to x_1.
Some elements of x_2 are positive and some negative, they characterize the two clusters.



matrices $m = \#$ edges $x_1 \ldots x_6$

$n = \#$ vertices

affinity matrix

$A_{ii} = e^{-d} \leftarrow$ distance

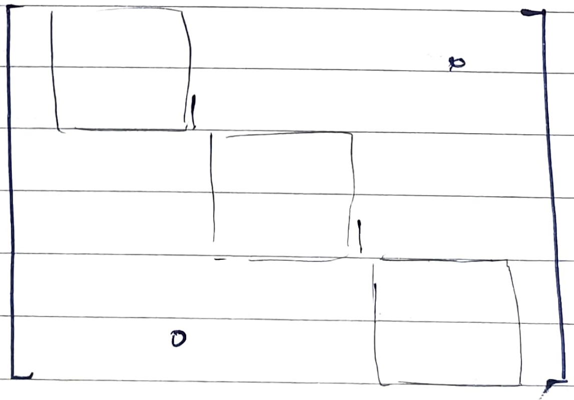
how to do clustering

A_{ij} = affinity matrix e^{-d}

$d_{ij} \to 0 \Rightarrow e^{ij} = 1$

$d_{ij} \to \infty \Rightarrow e^{ij} = 0$

A_{ij} =



max $\underline{\omega^T A \omega}$ st. $\|\omega\| = 1$

$\sum_i \sum_j \omega_i \omega_j A_{ij}$

$A_{ij} \to 0 \Rightarrow \omega_i \omega_j A_{ij} \to 1$

ω is the $\overline{EV}$ of A

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad \omega_i = 1 \quad \text{if} \quad > .$$

threshold the values.

adjacency matrix $\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ & & A & & & \end{bmatrix}$

$$A \begin{bmatrix} 1 \\ 1 \\ 1 \\ \\ 1 \end{bmatrix} = \begin{bmatrix} n-1 \\ n-1 \\ n-1 \\ \\ n-1 \end{bmatrix}$$

$$A \begin{bmatrix} 1 \\ 1 \\ 1 \\ \\ 1 \end{bmatrix} = (n-1) \begin{bmatrix} 1 \\ 1 \\ 1 \\ \\ 1 \end{bmatrix}$$

$n-1$ $\quad$ $n-1$

2n matrice

$$\begin{bmatrix} & A & \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} = (n-1) \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$