SVD, LEAST SQUARES AND LEAST NORM

$Ax = b$ → $b = 0$ system of homogeneous equations.

simple / fast → Opt prob. → EV

- Ch — Opt prob.
- LU — Least Sqr.
- QR — least norm.
- SVD : UDV^T

diagonal. (often sorted).

$$\underset{m \times n}{A} = \underset{m \times n}{U} \; \underset{n \times n}{D} \; \underset{n \times n}{V^T}$$

orthogonal $\quad U^TU = VV^T = V^TV = I$

$Ax = b$

$UDV^T x = b$

$\cancel{U^T U} D \underbrace{\cancel{V^T x}}_{\omega} = U^T b \to P$

$A = LU$

$Ax = b$

$LUx = b$

$L\omega = b$

$Ux = \omega$

(1) factorize A as UDV^T

(2) compute $P = U^T B$

(3) solve $D\omega = P$

(4) solve $V^T x = \omega \Rightarrow x = V\omega.$

(1) fact. A on LU

(2) solve $L\omega = b$

(3) solve $Ux = \omega.$

$$A^{-1} = (UDV^T)^{-1}$$
$$= VD^{-1}U^T$$

$$A^T = (UDV^T)^T$$
$$= VDU^T$$

$$A^TA = (UDV)^T UDV^T$$

$$= VDU^TUDV^T$$
$$A^TA = VD.DV^T$$
$$= VD^2V^T$$

$$A^TAV = \underbrace{VD^2}_{\text{diagonal}}$$

$$A^TAV_d = D^2V$$

$$A^TAV_1 = D_{11}^2V_1$$

V_1 is the eigen vector,
D_{11}^2 is eigen value of A^TA

$$A^TAV_i = D_{ii}^2V_i$$

Sorting D_{ii} and other matrices wont change A.

$A = UDV^T$

$\min_{A'} \|A - A'\|$

$A' = [\quad][\quad {}_{0}][\quad]$

$\text{rank}(A') = k.$

A' is the nearest ran -n-1

matrix to A

$A \qquad A^{-1}$

$(A + \lambda I)$ — regularization,

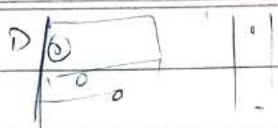
D_{ii} = eigen vectors (values) of A^TA, AA^T

$Ax = b$

$m \times n \quad n \times 1 \quad m \times 1$

$m > n$	$m < n$
↓	↓
more rows	more columns.
LSE	LN
MSE	
$\min \|Ax - b\|$	$\min \underline{\|x\|}$
	s.t.
$x = (A^TA)^{-1}A^Tb$	$(Ax = b)$
solve $\underbrace{A^TA}_{\downarrow\ \text{PSD}}x = A^Tb.$	$x = A^T(AA^T)^{-1}b.$

8463990548 Karthik



$$x^{*T}(x - x^*)$$

$$(A^T(AA^T)^{-1}b)^T (x - x^*)$$

$$((AA^T)^{-1}b)^T \underbrace{(A(x - x^*))}_{0}$$

L0
L1
L2

L0 - hard, sparse x.

$Ax = b$

$\min \|x\|$

LS

LN

$m > n$

$x^* = (A^TA)^{-1}A^Tb$

$m < n$

$x^* = A^T(AA^T)^{-1}b$

EV solution is an eigen vector.

"

$x^2y - \lambda(x^2y^2 - 1)$

$xy - \lambda(x+y-10)$

$xy - \lambda x - \lambda y + 10\lambda$

$\frac{\partial}{\partial x}$ $y - \lambda = 0$ — (1)

$\frac{\partial}{\partial y}$ $x - \lambda = 0$ — (2).

$\frac{\partial}{\partial \lambda}$ $-x - y + 10 = 0$ — (3).

$x = y = \lambda$

$\therefore 2\lambda = 10$

$\Rightarrow \lambda = 5$

$x = y = \lambda = 5.$

$x = y = 5.$

$-xy - \lambda(x+y-10)$

$-xy - \lambda x - \lambda y + 10\lambda$

$-y - \lambda = 0$

$-x - \lambda = 0$

$-x - y + 10 = 0$

$-\lambda + \lambda + 10 = 0$

$\lambda =$

$L = xy + \lambda(x+y-10)$

$= xy + \lambda x + \lambda y - 10\lambda$

$\frac{\partial}{\partial x}: \quad y + \lambda = 0$

$\frac{\partial}{\partial y}: \quad x + \lambda = 0$

$\frac{\partial}{\partial \lambda}: \quad x + y - 10 = 0$

$y = x = -\lambda$

$-2\lambda - 10 = 0$

$\Rightarrow \lambda = -10/2 = -5$

$\therefore x = y = -\lambda = 5.$

$x^2 + y^2 = 80$ farthest and closest from $(1,2)$

max $f(x,y) = (x-1)^2 + (y-1)^2$

St :

vector

$$L(x,\lambda) = x^T x + \lambda^T [Ax - b]$$

if you have multiple const. λ = vector.
connected together wrt. λ.

$$2x + y = 1$$

$$L(\lambda x) = x^T A^T A x - \lambda x^T x$$

$$\frac{\partial}{\partial x} = 0 \Rightarrow \boxed{A^T A x = \lambda x.}$$

$$\Downarrow$$

$$\Sigma x = \lambda x$$

matrix.

eigen value

eigen vector

$Ax = 0$

X trivial.

many solutions with a scale factor.
so let norm of $x = 1$.

constraint.

x is the eigen vector of $A^T A$.

$Ax = \lambda x$

λ^2 to be zero how you get smallest?

min $\|Ax\|$

same as min $\|Ax\|^2 \longrightarrow$ EV.

$A \Rightarrow UDV^T$

$A^T A \Rightarrow$.

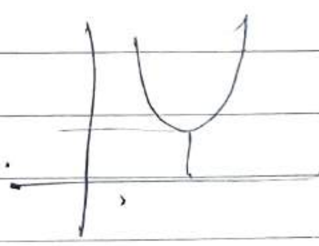
① 1st form A is non-singular & square

②. optimize least sqr. least norm.

closed form. soln.

global optimality.

convex optimization problem.



$x = A^{+} b$ ← psuedo inverse

A not full rank,

D_{ii} (SVD)

$D_{ii} < 0$.

$$\|Ax-b\| - \|x\|$$

$$(Ax-b)(Ax-b)^T - x^T x$$

$$AA^T x x^T \cancel{A} - Axb^T - A^T b x^T + bb^T - xx^T$$

$$\frac{\partial}{\partial x} \quad AA^T x - Ab^T - \cancel{A^T} bA^T - x = 0$$

$$\Rightarrow A\cancel{b^T}$$

$$\Rightarrow 2AA^T x - x = Ab^T + bA^T$$

$$x(AA^T - I) = Ab^T + bA^T$$

$$Ax = (AA^T - \cancel{A})^{-1}(Ab^T + bA^T)$$

1. ad
2. c
3. acd
4. bc
5. b.

LS soln. with cholesky.

$$x = (A^T A)^{-1} A^T b.$$

we want to compute as.

$$(A^T A)x = A^T b$$

$$Cx = D.$$

Cholesky

QR

SVD.

$$Ax_i = b_i$$

Many times to do.

1. costly facto.
2. easy compute

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

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

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

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

dont want to transmit basis, - DCT

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

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

eigen - standard trick, do cholesky of B

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

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

let $y = L^T y$

$x = L^{-T} y$

$\rightarrow A^{-1} y = \lambda y \leftarrow$ converted to classical problem.

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

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

$$A' y = \lambda y$$

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

$Ax = \lambda Bx$:

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

---

opt $x^T A x$ st. $\|x\| = 1$,

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

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

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

we know $\underline{\underline{x^T x = 1}}$

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

pick largest value,
so objective is λ.

min., pick smallest eigen value

$Ax = 0$

$\min \|Ax\|$ st. $\|x\|=1$

remove trival sols.
get a fixed magnitude.

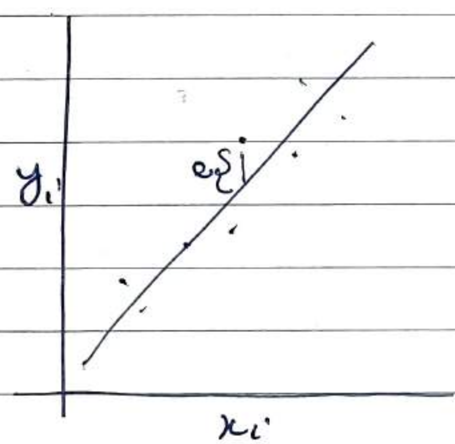
scalar.

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

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

$z^* = \lambda$.

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



$L2 \rightarrow$ LSE

$L1 \rightarrow$ LP

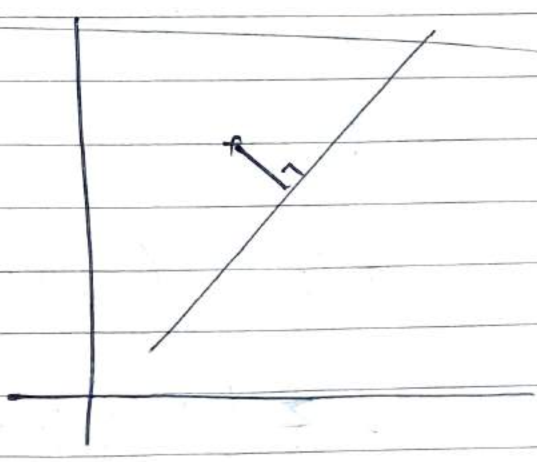
$L0 \rightarrow$ most points passing.

$L\infty \rightarrow$ distance to largest outlier will smaller.

---

suddenly becomes an EV problem.

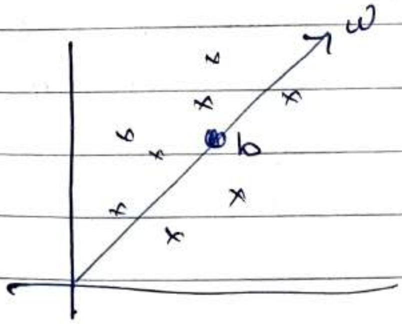
(eg. PCA).



2

variance is maximum.

$y_1 \dots y_N$



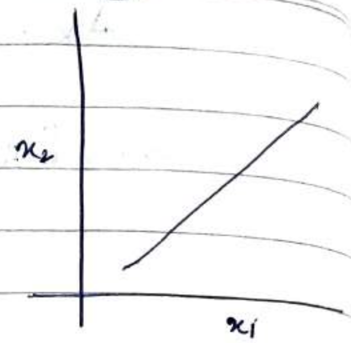
$x = w^T (y - b)$

↑ ↑

to find.

we would like variance to be maximized.

1



set of dimensions

$x_1 \dots x_n$

set of points

$x_1\ x_2\ x_3$

interested in finding the plane

no special axes.

orthogonal distance.

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

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

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

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

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

$$\underbrace{\frac{1}{N}\sum_i \omega^T (y_i - \bar{y})(y_i - \bar{y})^T \omega}$$

$$\omega^T \Sigma \omega - \lambda (1 - \omega^T \omega)$$ covariance of y_i

ω is the 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.

$$6\ (x-1)^2 + y^2 \qquad x+y=2$$ (b).

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

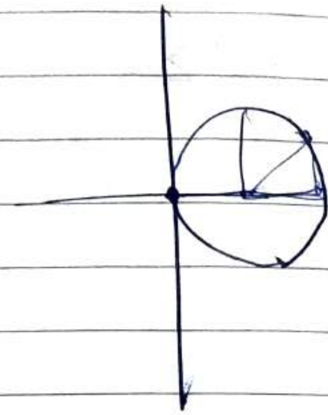
$$\frac{\partial L}{\partial x} = 0 \Rightarrow 2(x-1) - \lambda = 0 \quad 2(x-1) = \lambda$$

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

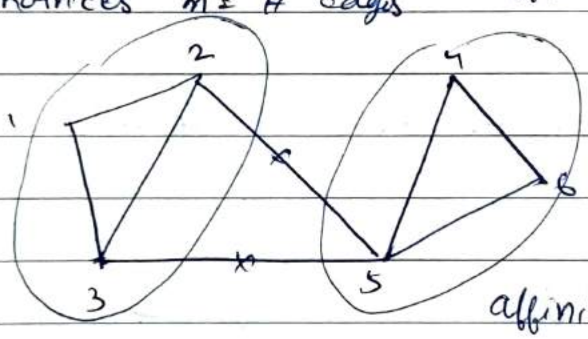
$(3/2, 1/2)$ $$\frac{\partial L}{\partial \lambda} = 0 \Rightarrow x+y=2$$

① b ② (a) (c) ③ a, b, c, d.

④ (a) (d)



matrices m = # edges n = # vertices.



$x_1 \dots x_6$

affinity matrix

$A_{ij} = e^{-d}$ → distance.

how to do clustering.

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

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

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

classmate

Date 28.2.2020
Page 7

$$A_{ij} = \begin{bmatrix} \boxed{} & & 0 \\ & \boxed{} & \\ 0 & & \boxed{0} \end{bmatrix}$$

max $\underline{\omega^T A \omega}$ $\sum_i \sum_j \omega_i \omega_j A_{ij}$

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

s.t. $\|\omega\| = 1.$

ω is the $e\bar{v}$ of A.

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

$\omega_i = 1$ of $>$

threshold the values.

adjacency matrix $\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ & & & A & & \end{bmatrix}$ $Ax \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} n-1 \\ \\ n-1 \end{bmatrix}$

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

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

2n vertices

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