Lecture Revision

11/08/24

Date
Page

Assignment / Quiz - 10
class Test - 20
midsem - 30
endsem - 40

11/08/24

linear algebra is the branch of mathematics concerning linear equations such as

$$a_1x_1 + \cdots + a_nx_n = b$$

→ in vector notation we can say $\underline{a}^T\underline{x} = b$
→ called a linear transformation of x.

Scalar

single number, represented in lower case italic x

Vector

array of number arranged in order
each no. identified by an index
Represented by lower case bold

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad n \times 1$$

we can think of vectors as points in space

Matrices

↳ 2-D array of numbers
↳ Denoted by bold typeface A

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}_{m \times n}$$ m rows, n column

Tensor

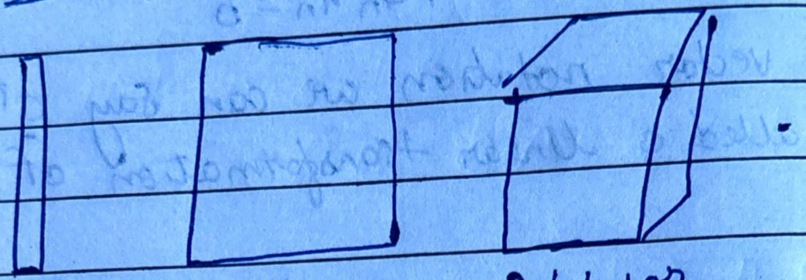
↳ Array with more than 2 dimensions

Ex:- RGB Color image has 3 axes

↳ denoted by bold type face. **A**

↳ element (i,j,k) of tensor denoted by $A_{i,j,k}$

Shape of tensors



1-d 2-d 3d tensor 4-d --5-d

vector metrices Tensor.

Dimension — No. of dependent variable

Transpose of matrix

$$(A^T)_{i,j} = A_{j,i}$$

→ interchange row and Column

→ mirror image across a diagonal line

Called main diagonal upper left to right corner

Ag- $A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$ $A^T = \begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{bmatrix}$

Date
Page

Vectors as special case of matrix

- Vectors are matrices with a single column.
- scalar is a matrix with one element

matrix addition

- we can add matrices to each other if they have same shape by adding corresponding elements

$$C = A+B : C_{ij} = A_{ij} + B_{ij}$$

Multiplying matrices

For product $C = AB$, A has to have same number of colums as the number of rows of B

$$C = AB \Rightarrow C_{ij} = \sum_k A_{ik} B_{kj}$$

element wise product is called Hadamard product

$$(A \odot B)_{ij} = A_{ij} \cdot B_{ij}$$

it is also called dot product.

Cosine similarity

$$\cos(\theta) = \frac{a \cdot b}{\|a\| \cdot \|b\|}$$

$a \cdot b$ is dot product

Date
Page

Matrices product properties

↳ Distributivity over addition

$$A(B+C) = AB+AC$$

↳ Associativity

$$A(BC) = (AB)C$$

↳ Not Commutative.

$A \cdot B = B \cdot A$ is not always true

↳ Dot product btw vectors is commutative

$$x^T y = y^T x$$

↳ Transpose of a matrix product has a simple form

$$(AB)^T = B^T A^T$$

Lecture Revision

Date 18/08/24
Page

→ Matrix inverse
→ Linear equations :- closed form solutions

Two closed form solutions
(1) matrix inversion $x = A^{-1}b$
(2) Gaussian elimination. → Time $O(n^3)$ nearly impossible for real world use case.

Practice Example

$$x + 3y - 2z = 5$$
$$3x + 5y + 6z = 7$$
$$2x + 4y + 3z = 8$$

→ Disadvantage of closed form solutions

→ How many solutions for $Ax = b$ exists ?

$$A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n = b_1$$
$$A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n = b_2$$
$$\vdots$$
$$A_{m1}x_1 + A_{m2}x_2 + \cdots A_{mn}x_n = b_m$$

System of equations with n variables and m equation.
Solution is :- $x = A^{-1}b$

Date
Page

we can get 3 solutions.

(a) unique solution.

$$2x + 3y = 9$$
$$9x + 2y = 11$$

we will get unique solution, because these are linear lines going to intersect at point

Slope :- $y = \frac{9-3x}{3}$

Slope = $\frac{dy}{dx}$

Here for both equation,
→ Slopes are different for y
→ intercept are different for y
∴ we will get unique solution

(b) infinite solution :-

$$2x + 3y = 4$$
$$4x + 6y = 8$$

we will get infinite solution because both line are same.
→ slope and intercept of y are same.
∴ we will get infinite solution

(c) no solution

$$2x+3y=4$$
$$2x+3y=9$$

we will get no solution
→ slope is same
→ intercept is different

in this case both lines will never converge.
∴ we will get no solution.

Span of a Set of vectors

span of a set of vectors: set of points obtained by a linear combination of those vectors.

→ A linear combination of vectors $\{v_1, v_2 -- v_n\}$ with coefficient c_i is

$$\sum c_i v_i$$

Ex:- Consider two vector in $\mathbb{R}^2$

$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

span is $\text{Span}(\bar{v}_1, \bar{v}_2) = \{c_1 v_1 + c_2 v_2 \mid c_1, c_2 \in \mathbb{R}\}$

span include all vectors formed by combination of v_1 and v_2

in this case $\text{Span}(v_1, v_2) = \mathbb{R}^2$ entire 2D space

$c_1 = 1, c_2 = 2$ $\quad 1 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$c_1 = -1, c_2 = 0$ $\quad -1\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$

Date
Page

Linear independence

now, if Linear combination is zero

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = 0$$

here vectors are linearly independent if and only if

$$c_1 = c_2 = c_3 \cdots = c_n = 0$$

Linear dependence

if anyone of coefficient is not zero then we can represent one in form of other then it is called as linearly dependent vectors

{ we are going to use this in finding orthogonal projection }

orthogonality of vectors

~ refers to condition when two vectors are at right angle 90° to each other. This happens if their dot product is zero

Two vectors u and v in $\mathbb{R}^n$ is orthogonal

if $u \cdot v = 0$

$$u \cdot v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

Geometrically :- if $u \cdot v = 0$

angle b/w two vectors satisfy.

$$\cos(\theta) = 0 \Rightarrow \theta = 90°$$

Ex:- $u = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ $v = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

$u \cdot v = 1 \cdot 1 + 0 \cdot 2 - 1 \cdot 1 = 0$

$\therefore$ $\bar{u}, \bar{v}$ are orthogonal.

Cosine Similarity

→ a measure of similarity b/w two vectors based on the cosine of the angle b/w them. it quantifies how aligned or similar two vectors are in terms of direction, regardless of their magnitude

Formula :-

Cosine Similarity $\cos(\theta) = \dfrac{u \cdot v}{||u|| \cdot ||v||}$

$u \cdot v$ is dot product

$||u|| = \sqrt{\sum u_i^2}$ $||v|| = \sqrt{\sum v_i^2}$ is magnitude

θ is angle b/w them.

Use of Vector in Regression

	# hours studied(x)	# hours playing games(y)	# classes missed(z)	Grade (o/p)
Student 1	10	3	0	87
Student 2	8	20	2	75
Student 2	5	1	5	63

This is regression problem.

Q) A student k 7hr study, 4hr games, missed 15 classes → Grade?

Norms

- used for measuring size of vector
- norms map vectors to non negative values
- norm of vector $x = \{x_1, \ldots x_n\}^T$ is distance from origin to x

- Norms ~~add~~ ~~act~~ as regularizer or as an constraint to size of vector to avoid overfitting

Types of norms

1) L_p Norm:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$$

Generalized form of L_1, L_2, and ∞ norms

$p=1$ L_1 norm

$p=2$ L_2 norm

$p=\infty$ infinity norm.

2) L_2 Norm (Euclidean Norm)

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} v_i^2}$$

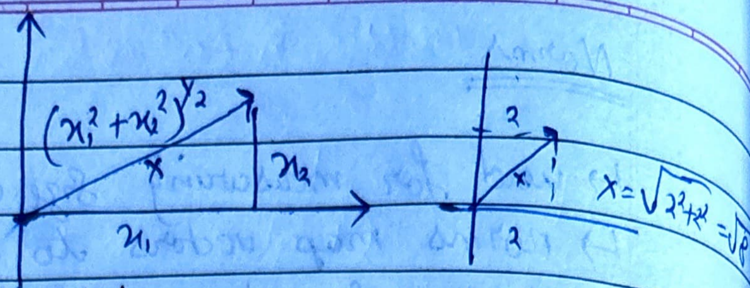
or Euclidean distance

measures the straight line distance from the origin to point represented by vector.

Ex:- $V = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$ $\|V\|_2 = \sqrt{1^2 + (-2)^2 + 3^2} = \sqrt{14}$

Date
Page

Its just pythagoras theorem



* squared euclidean norm is same as $x^T x$

3) L1 Norm (manhattan Norm)

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

For matrix

$$\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{m} |a_{ij}|$$

j = column i = row

measures manhattan distance, sum of absolute values of vector components

Ex:- $V = [1, -2, 3] \Rightarrow \|V\|_1 = |1| + |-2| + |3| = 6$

useful when 0 and non-zero have to be distinguished

Ex:- $0.1 \longleftrightarrow 0$ we can distinguish

taking square $(0.1)^2 \longleftrightarrow 0$ it become more closer to zero
0.01

L1 norm is used when there is sparsity. means number of non-zero element in vector is very less

$[0, 0.5, 0, 0, 0.1, 0]$

4) L^∞ norm (max norm)

$$\|x\|_\infty = \max_{i=1}^{n} |x_i|$$

Date
Page

Ex:- $x = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$ $\|x\|_\infty = \max(|1|, |-2|, |3|) = 3$

Ex:- $A = \begin{bmatrix} 7 & -2 & 2 & 2 \\ 2 & 8 & -2 & 3 \\ -3 & 2 & 11 & -4 \\ -2 & 3 & 2 & 10 \end{bmatrix}$

→ $\|A\|_\infty = ?$

$\Sigma |A_1| = 7+2+2+7 = 13$
$\Sigma |A_2| = 2+8+2+3 = 15$
$\Sigma |A_3| = 3+2+11+4 = 20$
$\Sigma |A_4| = 2+3+2+10 \; 17$

$\|A\|_\infty = \max(13, 15, 20, 17)$
$\|A\|_\infty = 20$

→ $\|A\|_1 = \cancel{[14, 15}$?

absolute column sum.

1st column $= |7| + |2| + |-3| + |-2| = 14$
2nd column $= |-2| + |8| + |2| + |3| = 15$
3rd column $= |2| + |-2| + |11| + |2| = 17$
4th column $= |2| + |3| + |-4| + |10| = 19$

$\|A\|_1 = \max(14, 15, 17, 19) = 19$

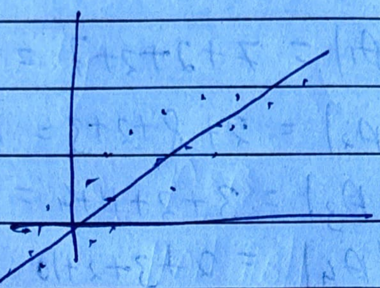
Date
Page

Use of norm in Regression

Let say we have a model (a mapping function btw input and output) given by linear regression.

x: is vector, w: weight vector.

$$y(x, w) = w_0 + w_1 x_1 + \dots + w_d x_d = w^T x$$



we need to find

y is line fitting to data.

w is one which we need to learn.

with non linear basis function ϕ_j (curve fitting)

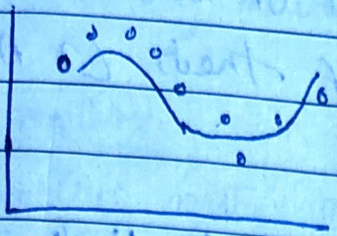
$$y(x, w) = w^T \phi(x)$$

↳ non linearity

Loss function

$$\tilde{E}(w) = \frac{1}{2} \sum_{n=1}^{N} \{ \underbrace{y(x_n, w)}_{\text{predicted}} - \underbrace{t_n}_{\text{Actual}} \}^2 + \frac{\lambda}{2} ||w||^2$$

Second term is a weighted norm called a regularized (to prevent overfitting) we need to minimize error b/w predicted and actual.



if we dont have ~~third~~ second term then ~~error~~ loss for error loss function will not converge.

LP Norm and Distance

- ↳ LP Norm form shape based on value of p
- ↳ L2 Norm or Euclidean norms yeild circle
- ↳ L1 Norm is rectangular or square
- ↳ L_∞ norm form straight line.

(8) Frobenius Norm (for matrices)

Size of matrix

$$||x||_F = \sqrt{\sum_{ij} a_{ij}^2}$$

$$A = \begin{bmatrix} 2 & -1 & 5 \\ 0 & 2 & 1 \\ 3 & 1 & 1 \end{bmatrix} \quad ||A|| = \sqrt{4+1+25+\dots+1}$$

$$= \sqrt{46}$$

Angle btw vectors

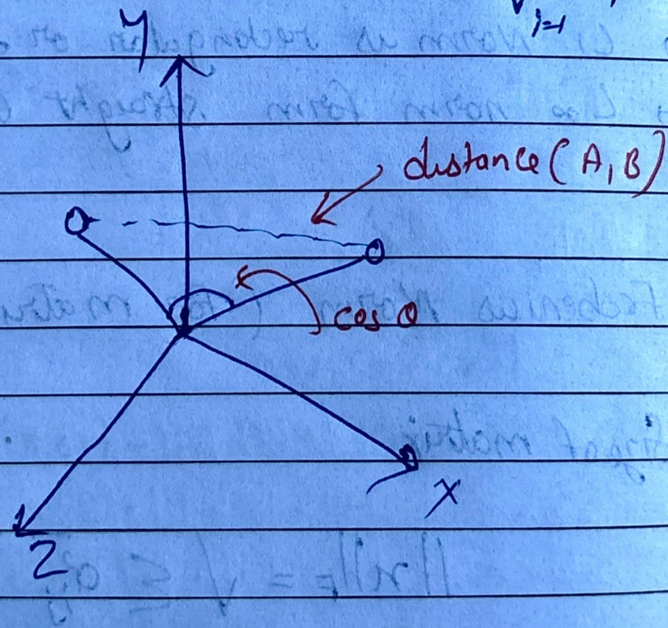
application of L2 norm

Dot product of two vectors can be written in terms of their L2 norms and angle θ btw them.

$$x^T y = \|x\|_2 \|y\|_2 \cos\theta$$

Cosine btw two vector is a measure of their similarity

$$\text{Similarity} = \cos\theta = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$



Date
Page

Diagonal matrix

- mostly zero, with non zero enteries only in diagonal
- used in inverting matrix

Symmetric matrix

- A symmetric matrix equals its transpose
  $$A = A^T$$
- accross diagonal, matrix is symmetry

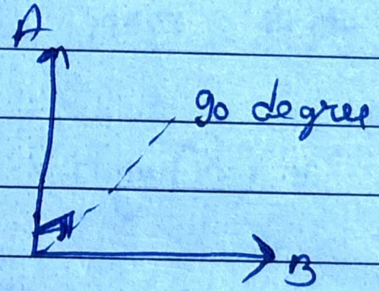
Special kind of vectors

- unit vector
  - A vector with unit norm $\|x\|_2 = 1$

- orthogonal vectors

$$x \cdot y = 0 \quad \text{or} \quad x^T y = 0$$

Vectors at 90° to each other, if vectors have non zero form



- orthonormal vectors

vector are orthogonal & have unit norm

Orthonormal matrix

$$\boxed{\begin{aligned} A^T A &= A A^T = I \\ \text{or} \quad A^{-1} &= A^T \end{aligned}}$$

their inverse is very easy to compute

25/08/24

Date
Page

Lecture Revision

Matrix decomposition

Matrices can be decomposed into factors to learn universal properties, just like integers

→ Decomposition of integer into prime factors
Form: $12 = 2 \times 2 \times 3$. we can discern that
↳ 12 is not divisible by 5
↳ any multiple of 12 is divisible by 3

↳ Any square matrix can be decomposed with Eigen value decomposition

↳ For non square matrix we can go with singular value decomposition

Eigen value decomposition

A is square matrix $n \times n$, eigen value decomposition

$$A = V \operatorname{diag}(\lambda) V^{-1}$$

V is eigen vectors
λ are eigen values

~~V is orthogonal i.e. $V^{-1} = V^T$~~

In eigen value decomposition V is not orthogonal

~~Ex~~

An eigen vector of a square matrix A is a non zero vector v such that multiplication by A only changes scale of v

$$Av = \lambda v$$

$v \rightarrow$ eigen vector

$\lambda \rightarrow$ eigen value.

$\rightarrow \lambda$ is scaling factor

if λ value is +ive then v is stretched

if λ value is -ive then v is compressed

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

This has non zero solution only if

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

On solving this we will get $\lambda_1, \lambda_2 \dots$ which are eigen values of A

Ex $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

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

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

$$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = 0$$

$$(2-\lambda)^2 - 1 = 0$$

$$4 - 2\lambda + \lambda^2 - 1 = 0$$

$$\lambda^2 - 2\lambda + 3 = 0$$

$$(\lambda - 1)(\lambda - 3) = 0$$

$$\lambda = 1, \quad \lambda = 3$$

For $\lambda = 1$

$$(A - I)V = 0$$

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

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$$v_1 + v_2 = 0$$

$$v_1 = 1 \quad v_2 = -1$$

$$\boxed{\text{eigen vector } V_{\lambda=1} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}}$$

For $\lambda = 3$

$$\begin{bmatrix} 2-3 & 1 \\ 1 & 2-3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$$-v_1 + v_2 = 0 \qquad v_1 = v_2 = 1$$

$$\boxed{\text{eigen vector } V_{\lambda=3} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}}$$

Date
Page

↳ Concatenate eigen vectors to form matrix V
↳ Concatenate eigen values to form vector λ

eigen decomposition of A

$$A = V \, \mathrm{diag}(\lambda) \, V^{-1}$$

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} +1/2 & +1/2 \\ +1/2 & -1/2 \end{bmatrix}$$

$$A = \begin{bmatrix} 1+0 & 0+3 \\ -1 & 0+3 \end{bmatrix} \begin{bmatrix} +1/2 & +1/2 \\ +1/2 & -1/2 \end{bmatrix}$$

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

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

$V_{\lambda=1}$ $V_{\lambda=3}$ eigen vectors are not orthogonal

For symmetric matrix

$A = A^T$

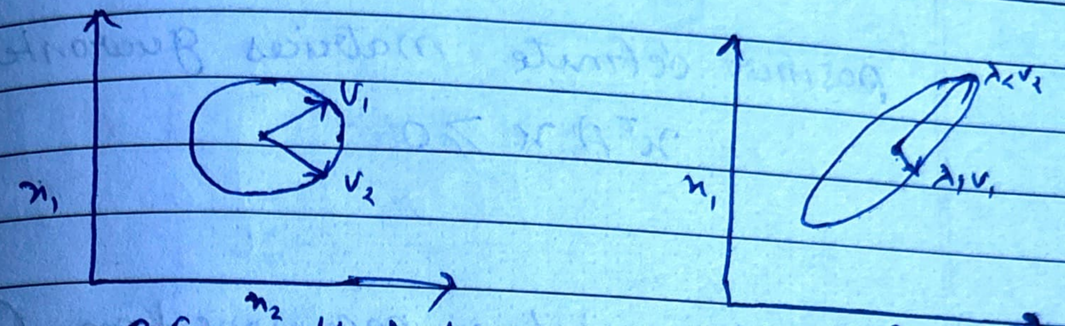
$$A = Q \lambda Q^T$$ ← here we need not to find inverse.

Q is orthogonal matrix composed of eigen vectors $A : \{ v^1, \ldots v^n \}$

Date
Page

Effect of Eigenvectors and Eigenvalues

when we multiply matrix by eigenvector, it is not going to change direction



Befor multipication   After multipication

→ Eigen decomposition is not unique when eigen values are same.

what does eigen decomposition tell us?

↳ matrix is singular if & only if eigenvalue is 0

what is importance of inverse ?
if we are not able to calculate matrix inverse then we will not be able to solve regression problem

→ Student, hr study, games problem see previous

⇒ if λ is zero than determinant $= 0$, we will not be able to find inverse

Date
Page

Positive Definati matrix

A matrix whose eigen values are all positive is called positive definite

positive definite matrices guarantee that

$$x^T A x \geq 0$$

Singular Value Decomposition (SVD)

SVD is applicable for both square and non square

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

orthogonal

U is upper triangular matrix
V is lower triangular matrix

SVD is more general than eigen decomposition
Here we do not need to calculate inverse

if A is $m \times n$, then U is $m \times m$, D is $m \times m$
V is $n \times n$.

Date
Page

Ex $A = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$

① $A^T A = \begin{bmatrix} 4 & 3 \\ 0 & -5 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix} = \begin{bmatrix} 25 & -15 \\ -15 & 25 \end{bmatrix}$

② $|A^T A - \lambda I| = 0$

$$\begin{vmatrix} 25-\lambda & -15 \\ -15 & 25-\lambda \end{vmatrix} = 0$$

$(25-\lambda)^2 - 225 = 0$

$625 - 50\lambda + \lambda^2 - 225 = 0$

$\lambda^2 - 50\lambda + 400 = 0$

$(\lambda - 40)(\lambda - 10) = 0$

$\lambda = 40, \lambda = 10$

③ $A^T A \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \lambda \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$

For $\lambda = 10$

$$\begin{bmatrix} 25 & -15 \\ -15 & 25 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 10 v_1 \\ 10 v_2 \end{bmatrix}$$

$25 v_1 - 15 v_2 = 10 v_1$

$-15 v_1 + 25 v_2 = 10 v_2$

$v_1 = v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Date
Page

L2 norm $= \sqrt{1^2+1^2} = \sqrt{2}$

Eigen Vector $V_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

$\lambda = 40$

$$\begin{bmatrix} 25 & -15 \\ -15 & 25 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 40v_1 \\ 40v_2 \end{bmatrix}$$

$25v_1 - 15v_2 = 40v_1$
$-15v_1 + 25v_2 = 40v_2$

$v_1 = -v_2 \quad v_2 = 1 \quad v_1 = -1$

L2 norm $= \sqrt{(1)^2 + (-1)^2} = \sqrt{2}$

Eigen vector $V_2 = \begin{bmatrix} -1/\sqrt{2} \\ +1/\sqrt{2} \end{bmatrix}$

→ Eigen values are always placed in desending order

(4) Find $\Sigma = \begin{bmatrix} \sqrt{40} & 0 \\ 0 & \sqrt{10} \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$

(5) $V = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$

Same sequence of λ

$\lambda = 40$ $\qquad \lambda = 10$

(6) find $U = AA^T$

$$AA^T = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 0 & -5 \end{bmatrix}$$

$$= \begin{bmatrix} 16 & 12 \\ 12 & 34 \end{bmatrix}$$

$$|AA^T - \lambda I| = 0$$

$$\begin{vmatrix} 16-\lambda & 12 \\ 12 & 34-\lambda \end{vmatrix} = 0$$

$(16-\lambda)(34-\lambda) - 144 = 0$

$544 - 16\lambda - 34\lambda + \lambda^2 - 144 = 0$

$400 - 50\lambda + \lambda^2 = 0$

$\boxed{\lambda = 40,\ 10} \longrightarrow$ This has to be same as A^TA

checkpoint

for $\lambda = 10$

$$AA^T \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \lambda \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\begin{bmatrix} 16 & 12 \\ 12 & 34 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 10 \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$16v_1 + 12v_2 = 10v_1$

$12v_1 + 34v_2 = 10v_2$

$\cancel{16}v_1 = -\cancel{24}^{2}v_2 \qquad v_1 = -2v_2$

$\cancel{v_1 = 1} \quad \cancel{v_2 = -2} \quad v_2 = 1 \quad v_1 = -2$