Lecture 1 : Review

Data can be represented as a matrix.

Data can be rep. as a mtx

Ex. Data mtx

meas. height, weight, age on $N=4$

$$X = \begin{array}{c} \begin{bmatrix} 6.1 & 100 & 10 \\ 5.5 & 150 & 20 \\ 7.3 & 200 & 25 \\ 6 & 250 & 75 \end{bmatrix} \\ \begin{array}{ccc} \text{height} & \text{weight} & \text{age} \end{array} \end{array}$$

So we have a $N \times P$ $(P=3)$ data mtx

View a mtx as a collection of rows

$$X = \begin{bmatrix} — & x_1 & — \\ — & x_2 & — \\ & \vdots & \\ — & x_N & — \end{bmatrix}$$

$x_n \in \mathbb{R}^P$ is an observation

Ex. $x_1 = (6, 100, 10)$

or a collection of cols

$$X = \begin{bmatrix} | & | & & | \\ X_1 & X_2 & \cdots & X_p \\ | & | & & | \end{bmatrix}$$

where $x_p \in \mathbb{R}^N$ is a variable

Ex. $X_1 = (6.1, 5.5, 7.3, 6)$ is height var.

Inner Products:

If $a, b \in \mathbb{R}^p$ then the inner prod is

$$a \cdot b = a'b = a^T b = \sum_{k=1}^{p} a_k b_k$$

Norm: The norm/length

$$\|a\| = \sqrt{\sum_{k=1}^{p} a_k^2} = \sqrt{a^T a}$$

What about matrices?

$A \in \mathbb{R}^{m \times n}$; $B \in \mathbb{R}^{n \times p}$ then $AB \in \mathbb{R}^{m \times p}$

must match

4 ways to defn AB:

Inner product

① $$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

$$= \text{row } i \text{ of } A \cdot \text{col } j \text{ of } B$$

Linear Combination

② LC of Cols of A

$$B = \begin{bmatrix} | & | & & | \\ B_1 & B_2 & \cdots & B_p \\ | & | & & | \end{bmatrix} \text{ then } AB = \begin{bmatrix} | & | & & | \\ AB_1 & AB_2 & \cdots & AB_p \\ | & | & & | \end{bmatrix}$$

note $AB_k = $ LC of cols of A

(3) LC of Rows of B

LC of rows of B

$$A = \begin{bmatrix} - & a_1 & - \\ - & a_2 & - \\ & \vdots & \\ - & a_m & - \end{bmatrix} \quad \text{then} \quad AB = \begin{bmatrix} - & a_1 B & - \\ - & a_2 B & - \\ & \vdots & \\ - & a_m B & - \end{bmatrix}$$

(4) Sum of OPs : $a^T b = $ inner prod ← a number

Outer Products

$ab^T = $ outer product

pxp mtx

$$AB = \sum_{k=1}^{n} \overset{m\times1}{a_k} \overset{1\times p}{b_k^T} \leftarrow m \times p$$

col of A   row of B

Matrix Norm?

Vector : $\|a\| = \sqrt{a^T a} = \sqrt{\sum_{k=1}^{p} a_k^2}$

Mtx : $A \in \mathbb{R}^{m \times n}$

$$\|A\|_F = \sqrt{\sum_i \sum_j A_{ij}^2} = \sqrt{tr(A^T A)}$$

$$= \sqrt{tr(AA^T)}$$

Frobenius

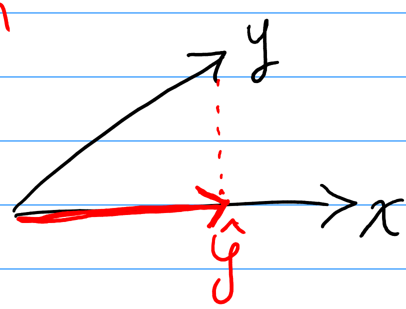
Projection:

$x, y \in \mathbb{R}^P$ then the proj of y onto x

is

u_x unit vec. in direction of x

$$\hat{y} = u_x u_x^T y$$



$$u_x = \frac{x}{\|x\|}$$

$$= \frac{x}{\|x\|} \frac{x^T}{\|x\|} y$$

$$\|x\| = \sqrt{x^T x}$$

$$= \frac{x x^T}{\|x\|^2} y$$

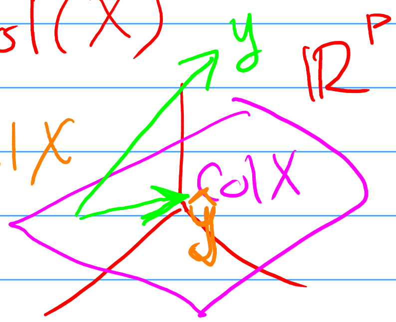
$$= \frac{x x^T y}{x^T x}$$

$$\boxed{\hat{y} = x(x^T x)^{-1} x^T y}$$

What about a mtx? X $\quad$ Col(X)

$\hat{y}$ = proj of y onto col X

= closest vec. to y in Col X



$$\boxed{\hat{y} = X(X^T X)^{-1} X^T y}$$

Orthogonality:

Unit: $\|u\| = 1$

both = ortho-normal

Orthogonal: $u^T v = 0$

Orthogonal Mtx: $\Phi \in \mathbb{R}^{N \times N}$ is called orthogonal if cols of Φ are mutually ortho-normal

1. cols are unit vectors
2. Cols are orthogonal

Can show

$$\Phi^T \Phi = I = \Phi \Phi^T$$

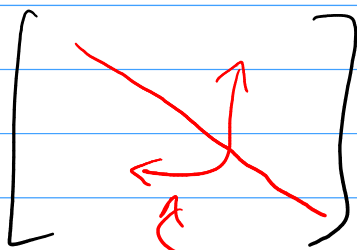
i.e. $\Phi^{-1} = \Phi^T$.

Eigen-Vectors/values

If A a mtx then v is an e-vector assoc. w/ e-val. λ if

$$Av = \lambda v$$

If A symmetric matrix

$A =$ 

symetric over main diag

e.g. $A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ $\qquad$ $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 1 \end{bmatrix}$

Eigen-value Decomp (EVD)

If A is symmetric then $\quad$ square

$$\boxed{A = Q D Q^T}$$

$N \times N$

where

$$Q = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_N \\ | & | & & | \end{bmatrix}$$

orthog mtx w/ cols that are e-vectors of A

$$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{bmatrix}$$

corresp. e-vals

Can generalize to SVD

Singular Value Decomp.

If I have any mtx $A \in \mathbb{R}^{m \times n}$

then

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

① U is orthog and cols are e-vecs of $\underset{m\times m}{AA^T}$

② V is orthog and cols ar e-vecs of $\underset{n\times n}{A^TA}$

③ $D = \left[\begin{array}{c|c} \sigma_1 \cdots \sigma_r & 0 \\ \hline 0 & 0 \end{array}\right]$ $r = \text{rank}(A)$

$\sigma_i = \sqrt{\lambda_i}$ where λ_i is e-val of A^TA or AA^T.