Lecture 16

let $x \in \mathbb{R}^N$ ← a variable (col of X)

assume $\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n = 0$

then $var(x) = \frac{1}{N-1}\sum_{n=1}^{N}(x_n - \bar{x})^2 = \frac{1}{N-1}\sum_n x_n^2$

$\propto \sum_n x_n^2 = x^T x$

Similarly if $y \in \mathbb{R}^N$ is some other var and either $\bar{x} = 0$ or $\bar{y} = 0$ then

$Cor(x,y) \propto Cov(x,y) \propto x^T y$

So uncorrelated $\approx$ orthogonal

Similarly if X is $N \times P$ data mtx (assume cols are mean-centered)

$Cov(X)$ is a mtx $(P \times P)$ where

$Cov(X)_{ij} = cov(X_i, X_j)$

$Cov(X)_{ii} = cov(X_i, X_i) = Var(X_i)$

Turns out $\text{Cov}(X) \propto X^T X$

PCA: $Z = XW$

Want: 1 maximize diag elements of $\text{Cov}(Z)$

(max sum of diag elemts $\text{Cov}(Z)$)

2 off diag elements of $\text{Cov}(Z)$ to be zero

3 cols of W are unit vecters

$$\text{Cov}(Z) \propto Z^T Z \qquad \left(\text{Cov}(Z) = \frac{1}{N-1} Z^T Z\right)$$

$$= (XW)^T (XW)$$

$$= W^T X^T X W$$

How to find W?

$$X = UDV^T$$

then $$W^T X^T X W = W^T V D^T \overbrace{U^T U}^{I} D V^T W$$

$$= W^T V D^T D V^T W$$

$$D = \left[\begin{array}{c|c} D_* & 0 \\ \hline 0 & 0 \end{array}\right] \qquad D_* = \operatorname{diag}(\sigma_i)$$

$r \times r$ mtx, $r = \operatorname{rank}(X)$

$$D^T D = \left[\begin{array}{c|c} D_*^2 & 0 \\ \hline 0 & 0 \end{array}\right]$$

diag

Consider $W^T V \overbrace{D^T D} V^T W$

How to make this diag?

Let $W = V$ then this is $V^T V D^T D V^T V$

$$= D^T D$$

Want W to be $p \times q$ and V is $p \times p$

Want sum of diag elements to be as large as possible

I can do this if I choose W as the first q cols of V

b/c then sum of diag. elements is

$$\operatorname{tr}(D^T D) = \sum_{j=1}^{q} \sigma_j^2$$

Punchline: PCA

1) make X by mean-centering vars

2) $X = UDV^T$

3) $W = V_q$ ← first q cols of V

4) $Z = XW = XV_q$

Comments: 1) $z_i = X w_i$

$= X v_i$

$= \sigma_i u_i$

$X v_i = \sigma_i u_i$

notice! $\mathrm{Var}(z_i) \propto z_i^T z = (\sigma_i u_i)^T (\sigma_i u_i)$

$= \sigma_i^2 u_i^T u_i$

$= \sigma_i^2$

$\mathrm{Var}(z_i) = \frac{1}{N-1} \sigma_i^2$

2)

Total Var of $z_i = \sum_{i=1}^{q} \mathrm{Var}(z_i) = \frac{1}{N-1} \sum_{i=1}^{q} \sigma_i^2$

Pct. of Var Captured:

$$= \frac{\cancel{\frac{1}{N-1}} \sum_{i=1}^{q} \sigma_i^2}{\cancel{\frac{1}{N-1}} \sum_{i=1}^{p} \sigma_i^2}$$

tot. var of PCs

tot. var of orig. data.

③ $z_i = \sigma_i u_i$ so $z_i \propto u_i$

kinda optional

④ mean-centering X

kinda optional

If I don't then $z_1 \approx$ mean of vars

<u>Could consider:</u>

$$\underbrace{X_q}_{N \times P} = X P_W$$

$WW^T = V_q V_q^T$

$$= \underbrace{X V_q}_{\substack{Z \\ N \times q}} \underbrace{V_q^T}_{\text{basis}}$$

$$\doteq U D \underbrace{V^T V_q}_{\left[\begin{smallmatrix} I \\ 0 \end{smallmatrix}\right]} V_q^T$$

first q cols U

$$= U_q D_q V_q^T$$

$q \times q$ diag

truncated SVD

X_q has rank $= q$

<u>Theorem: Eckart-Young Theorem</u>

X_q is the best rank-q approx of X.

$$X_q = \underset{B:\, \text{rank}(B)=q}{\text{argmin}} \; \|X - B\|$$