PCA:

Assume cols of X have been mean-centered

$$\hat{Cov}(Z) \propto Z^T Z$$

if $Z = XW$ then

$$\hat{Cov}(Z) \propto (XW)^T XW$$

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

Goal: $\hat{Cov}(Z)$ to be diag, with large elements on diag

How do we find W?

Consider $X = UDV^T$

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

$$= W^T V \underbrace{D^T D}\, V^T W$$

$$D = \left[\frac{D_* \mid 0}{} \right]$$

How do we make diag?

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

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

Consider $W = V$ then $W^T V = V^T W = I$

and then

$$\hat{Cov}(z) = D^T D$$

Problem: - want W to be $p \times q$ mtx and

V is a $P \times P$

- Want the sum of diag elements to be as large as possible.

Soln: I can do this if I let W be the first q cols of V

$$W = \underline{V_{1:q}}$$

$$\underbrace{\qquad}_{p \times q}$$

then $W = V_{1:q}$ then

$$\hat{Cov}(z) = W^T V D^T D V^T W$$

$$= \underbrace{V_{1:q}^T}_{\begin{bmatrix} I_q \\ \hline 0 \end{bmatrix}} V D^T D V^T V_{1:q}$$

$$= D_{*\,1:q} = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_q^2 \end{bmatrix}$$

So then sum of diag

$$\sum_{j=1}^{q} \sigma_j^2 .$$

Pipeline: PCA

① center cols of X

(0) mean center cols of X

(1) Calc SVD of X as $X = UDV^T$

(2) $W = V_{1:q}$

(3) $Z = XW = XV_{1:q}$

Comments:

$$\boxed{X v_i = \sigma_i u_i}$$

(1) $z_i = X w_i$

$= X v_i$

$= \sigma_i u_i$

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

$= \sigma_i^2 u_i^T u_i$

$= \sigma_i^2$

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

(2) Total var of Z

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

Pct. of var. captured by first q PCs

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

← amt. cap. var.

← total var.

(3) $z_i = \sigma_i u_i$

Kinda optional $z_i \propto u_i$

(4) We first mean-centred X

Kinda optional

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

Could consider:

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

proj. onto $Col(W)$

$= W W^T$

$= V_{1:q} V_{1:q}^T$

$$= \underbrace{X V_{1:q}}_{\substack{Z \\ \text{coords in } q\text{-dim'l subspace}}}^{N \times q} \underbrace{V_{1:q}^T}_{\text{basis}}$$

$P_A = A(A^T A)^{-1} A^T$

Cols of A orthog.

$A^T A = I$, $AA^T \neq I$

$= AA^T$

$$= UDV^T \underbrace{V_{1:q}}_{\left[\begin{smallmatrix} I_q \\ 0 \end{smallmatrix}\right]} V_{1:q}^T$$

$q \times q$ diag upper

$$\underbrace{X_q}_{N \times P} = U_{1:q} D_q V_{1:q}^T \quad \text{(truncated SVD)}$$

first q cols of U — first q cols of V

X_q has rank q.

X_q has rank q.

Theorem: **Eckart-Young Theorem**

X_q is the best rank-q approx of X:

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

$$\|A\|^2 = \sum_{i,j} A_{ij}^2$$
