FYS-STK 4155, NOV 29, 2022

PCA :

$$X \in \mathbb{R}^{n \times p}$$

$$X = \begin{bmatrix} x_{00} & x_{01} \\ x_{10} & x_{11} \\ x_{20} & x_{21} \end{bmatrix} = \begin{bmatrix} | & | \\ x_0 & x_1 \\ | & | \end{bmatrix}$$

$$\mathrm{cov}[x_i, x_j] = \frac{1}{n} \sum_k (x_{ki} - \mu_i)(x_{kj} - \mu_j)$$

$$\mathrm{cov}[x_0, x_1] = \frac{1}{n} \left(\bar{x}_{00} \bar{x}_{01} + \bar{x}_{10} \bar{x}_{11} + \bar{x}_{20} \times \bar{x}_{21} \right)$$

$$\left(\bar{x}_{ij} = x_{ij} - \mu_j \right)$$

$$X \to \bar{X} = \begin{bmatrix} \bar{x}_{00} & \bar{x}_{01} \\ \bar{x}_{10} & \bar{x}_{11} \\ \bar{x}_{20} & \bar{x}_{21} \end{bmatrix}$$

$$\mathrm{cov}[\bar{x}_0, \bar{x}_1] = \bar{x}_0^T \bar{x}_1$$

Define covariance matrix of design matrix

$$C[\mathbf{X}] = \frac{1}{n} \bar{X}^T \bar{X}$$

$$\bar{X} \in \mathbb{R}^{3 \times 2} \Rightarrow C[X] \in \mathbb{R}^{2 \times 2}$$

$$C[x] = \begin{bmatrix} cov[x_0 x_0] & cov[x_0 x_1] \\ cov[x_1 x_0] & cov[x_1 x_1] \end{bmatrix}$$

$$cov[x_0 x_0] = \frac{1}{n} \sum_{k=0}^{n-1} (x_{k0} - \mu_0)^2$$

$$\left(\mu_0 = \frac{1}{n} \sum_{k=0}^{n-1} x_{k0} \right)$$

$$= \sigma_0^2$$

$$C[x] = \begin{bmatrix} \sigma_0^2 & cov[x_0 x_1] \\ cov[x_1 x_0] & \sigma_1^2 \end{bmatrix}$$

Singular value decomp (SVD)

$$X = U \Sigma V^T \quad (\in \mathbb{R}^{n \times p})$$

$$U U^T = U^T U = \mathbb{1} \qquad U \in \mathbb{R}^{n \times n}$$

$$V V^T = V^T V = \mathbb{1} \qquad V \in \mathbb{R}^{p \times p}$$

$$\Sigma = \begin{bmatrix} \sigma_0 & & & & \\ & \sigma_1 & & & \\ & & \sigma_2 & & \\ & & & \ddots & \\ & & & & \sigma_{p-1} \\ & & & & 0 \end{bmatrix}$$

$$\Sigma \in \mathbb{R}^{n \times p}$$

singular values $\sigma_0 > \sigma_1 > \dots$
$> \sigma_{p-1} > 0$

$$X^T X = V \Sigma^T U U^T \Sigma V^T$$

$$= V \Sigma^T \Sigma V^T \in \mathbb{R}^{p \times p}$$

$$X^T X V = V \Sigma^T \Sigma = V \Sigma^2$$

$$= V \begin{bmatrix} \sigma_0^2 & & & \\ & \sigma_1^2 & & \\ & & \ddots & \\ & & & \sigma_{p-1}^2 \end{bmatrix}$$

$$V = \begin{bmatrix} | & | & & | \\ v_0 & v_1 & \cdots & v_{p-1} \\ | & | & & | \end{bmatrix}$$

$$v_i^T v_j = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & \text{else} \end{cases}$$

$$(X^T X) v_i = v_i \sigma_i^2$$

eigenvalues of $X^T X$ are given by $\Sigma^T \Sigma = \Sigma^2$ and has eigenvectors v_i

$$C[X] = \frac{1}{n} X^T X \Rightarrow$$

eigenvalues of $C[X]$ are given by σ_i^2 / n

$$\mathbb{E}\left[\frac{1}{n} x^T x\right] = C[x]$$

To find eigenvalues of $C[x]$, we perform a transformation with an orthogonal/unitary matrix S, $SS^T = S^T S = \mathbb{1}$

$$S\,C[x]\,S^T = D = C[y]$$

$$\mathbb{E}[S x^T x S^T] = S\,\mathbb{E}[x^T x]\,S^T$$

$$= S\,C[x]\,S^T = C[y]$$

$$= D = \begin{bmatrix} \lambda_0 & & 0 \\ & \ddots & \\ 0 & & \lambda_{p-1} \end{bmatrix}$$

$C[y]$ is decorrelated, No nondiagonal elements.

λ_0 is variance of $\mathrm{var}[y_0]$

With SVD, we have

$$\mathrm{var}[y_i] = \sigma_i^2/n$$

$$\text{Total variance} = \sum_{i=0}^{p-1} \sigma_i^2/n$$

Principal component analysis approximates the total variance by simply leaving terms with small singular values

Max variance with one component only.

vector s_0 $\left[V = \begin{bmatrix} | & | & & | \\ v_0 & v_1 & \cdots & v_{p-1} \\ | & | & & | \end{bmatrix} \right.$

$s_0^T s_0 = \underline{1}$ $\quad s_0 \in \mathbb{R}^p$

Maximize variance with the above constraint

$$\mathcal{L} = s_0^T C[x] s_0 + \lambda_0 (1 - s_0^T s_0)$$

calculate derivatives wrt s_0^T and λ_0

$$C[x] s_0 = \lambda_0 s_0$$

$$s_0^T s_0 = \underline{1}$$

s_0 is an eigenvector of $C[x]$ with eigenvalue λ_0

$$s_0^T C[x] s_0 = C[y] = \lambda_0$$

$$= var[y_0]$$

with more eigenvectors

$$s_1^T s_0 = 0 \qquad s_1^T s_{\underline{1}} = \underline{1}$$

$$\mathcal{L} = s_1^T C[x] s_1 + \lambda_1 (1 - s_1^T s_1)$$
$$+ \gamma s_1^T s_0$$

Take ~~derivatives~~ wrt s_1^T, λ_1, γ

$\rightarrow s_1$: $\quad C[x]\, s_1 + \gamma/2\, s_0 = \lambda_1 s_1$

λ : $\quad s_1^T s_1 = 1 \qquad \gamma$: $\quad s_1^T s_0 = 0$

$$s_0^T \Big[\quad \underbrace{s_0^T C[x]}_{\lambda_0} s_1 + \frac{\gamma}{2} \cdot 1 = \lambda_1 \underbrace{s_0^T s_1}$$

$$\gamma = (\lambda_1 - \lambda_0) \underbrace{s_0^T s_1}_{=0} = 0$$

$\Rightarrow$

$$C[x]\, s_1 = \lambda_1 s_1$$

s_1 is an eigenvector of $C[x]$, and λ_1 is an eigenvalue.

$$\lambda_1 = var[y_1]$$

for all eigenvalues λ_i, and s_i,

$$\mathcal{L} = s_i^T C[x]\, s_i + \lambda_i (1 - s_i^T s_i) + \sum_{j=0}^{i-1} \gamma_j\, s_i^T s_j$$

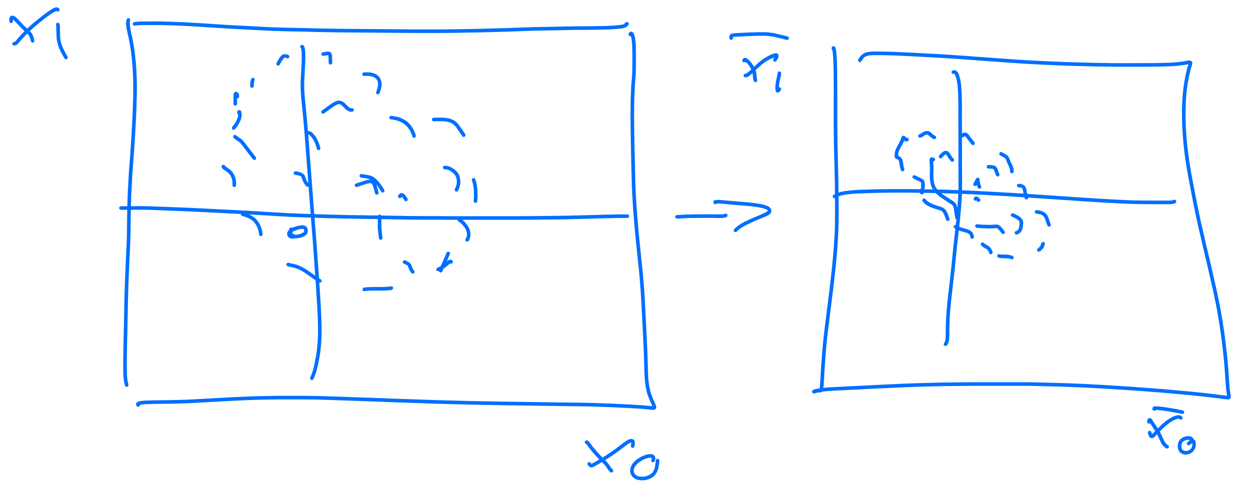
Basic steps

$$X = \begin{bmatrix} x_{00} & x_{01} \\ x_{10} & x_{11} \\ \vdots & \vdots \\ x_{n-1\,0} & x_{n-1\,1} \end{bmatrix} = \begin{bmatrix} | & | \\ x_0 & x_1 \\ | & | \end{bmatrix}$$

(i) mean value subtraction

$$X \rightarrow \bar{X} = X - \mu$$

$$\mu = \begin{bmatrix} | & | \\ \mu_0 & \mu_1 \\ | & | \end{bmatrix}$$

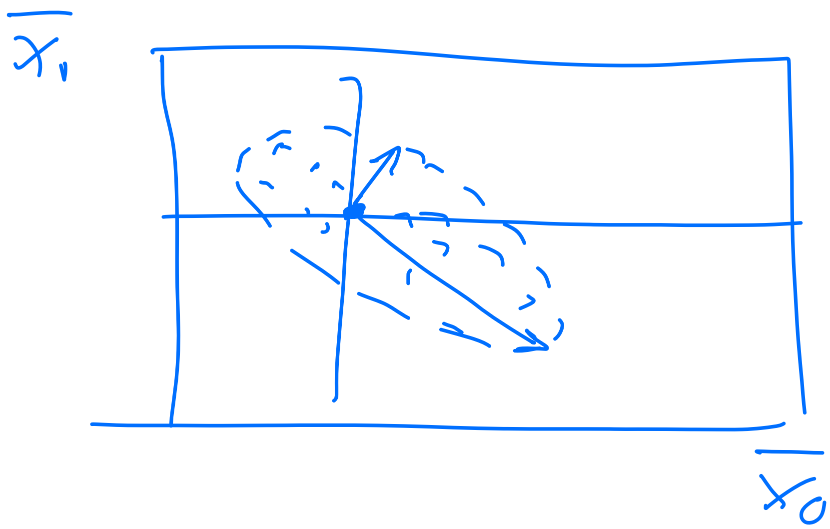


(ii) standardization. Divide the points with the standard deviation of a data set

Gives dimensionless quantities (and no units)

3) Eigendecomposition of the covariance matrix

Eigenvectors are scaled by the magnitude of the corresponding eigenvalues.



4) Projection step: project any data point onto the principal subspace