Lecture October 23

Principal component analysis (PCA): Favorite dim-Reduction algorithm.

Fitting a low-dimensional subspace of dim $d \ll D$ to a set of (multivariate) points $\{x_1 \dots x_N\}$ in a high dim space $\mathbb{R}^D$

covariance matrix $\mathbb{C}$

$x \in \mathbb{R}^n$

$y \in \mathbb{R}^n$

$$\mathbb{C}[x,y] = \begin{bmatrix} \mathrm{cov}[x,x] & \mathrm{cov}[x,y] \\ \mathrm{cov}[y,x] & \mathrm{cov}[y,y] \end{bmatrix}$$

$$\mathrm{cov}[x,y] = \frac{1}{n} \sum_{i=0}^{n-1} (x_i - \bar{x})(y_i - \bar{y})$$

$$\mathrm{var}[x] = \frac{1}{n} \sum_{i=0}^{n-1} (x_i - \bar{x})^2$$

$$\mathbb{C}[\underline{x,y}] = \begin{bmatrix} \mathrm{var}[x] & \mathrm{cov}[x,y] \\ \mathrm{cov}[y,x] & \mathrm{var}[y] \end{bmatrix}$$

$$\mathrm{corr}[x,y] = \frac{\mathrm{cov}[x,y]}{\sqrt{\mathrm{var}[x]\,\mathrm{var}[y]}}$$

correlation matrix $K[x, y]$

$$K[x, y] = \begin{bmatrix} 1 & corr[x, y] \\ corr[x, y] & 1 \end{bmatrix}$$

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

$$X = \begin{bmatrix} x_{00} & x_{01} & \cdots & x_{0\,p-1} \\ x_{10} & & & \\ \vdots & & & \\ x_{n-1\,0} & - & - & x_{n-1\,p-1} \end{bmatrix}$$

$$X = [\mathbf{x}_0 \ \mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_{p-1}]$$

Note : many texts define
$X \in \mathbb{R}^{p \times n}$

$$C[X] = \begin{bmatrix} var[x_0] & cov[x_0, x_1] & \cdots & cov[x_0, x_{p-1}] \\ cov[x_1, x_0] & var[x_1] & & \\ & & \ddots & \\ cov[x_{p-1}, x_0] & & & var[x_{p-1}] \end{bmatrix}$$

$C \in \mathbb{R}^{p \times p}$

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

$$X \in \mathbb{R}^{p \times n} \qquad C[X] = \frac{1}{n} X X^T$$

$$X = \begin{bmatrix} x_{00} & x_{01} \\ x_{10} & x_{11} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_0 & \mathbf{x}_1 \end{bmatrix}$$

$$\mathbb{E}[X^T X] = \frac{1}{n} X^T X = \frac{1}{n} \begin{bmatrix} x_{00}^2 + x_{01}^2 & x_{00}x_{10} + x_{01}x_{11} \\ x_{10}x_{00} + x_{11}x_{01} & x_{10}^2 + x_{11}^2 \end{bmatrix}$$

$$= \begin{bmatrix} \mathrm{var}[x_0] & \mathrm{cov}[x_0, x_1] \\ \mathrm{cov}[x_1, x_0] & \mathrm{var}[x_1] \end{bmatrix}$$

$$= C[X]$$

X = Feature matrix.

$n \gg p$: seldom to perform PCA

$n \simeq p$: PCA of interest if $p > 10^2$

$p \gg n$ PCA is normally applied.

Toward the PCA theorem

$$C[X] = \mathbb{E}[X^T X]$$

(recall OLS/Ridge
$$\mathrm{var}[\beta] \propto (X^T X)^{-1})$$

orthogonal transformation

$$S;\quad S^T S = S S^T = \mathbb{1} \qquad S^T = S^{-1}$$

$$S = [s_0, s_1, \dots s_{p-1}]$$

$$s_i \in \mathbb{R}^p \qquad s_i^T s_j = \delta_{ij}$$

$S^T C[x] S = C[y]$ has diagonal elements

$$[\lambda_0, \lambda_1, \dots \lambda_{p-1}]$$

$$C[y] = \mathbb{E}[(Sx)^2]$$

X has been normalized by subtracting the mean value.

$$S: \quad S^T C[X] S = S^T \mathbb{E}[X^T x] S$$

$$s \cdot / C[y] = s' \mathbb{E}[x'x] s$$

$$s \cdot C[y] = \underline{\mathbb{E}[x'x]} s \qquad \overbrace{}^{\frac{1}{n} x^T x} \in \mathbb{R}^{p \times p}$$

$\underline{C[y]}$ is diagonal, $\lambda_0 > \lambda_1 > \lambda_2 \ldots \lambda_{p-1}$

$$s_i \lambda_i = C[X] s_i$$

s_i are eigenvectors of $C[X]$ and λ_i are the eigenvalues.

orthogonal transformation

$S \in \mathbb{R}^{2 \times 2}$

$$S = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

$$\boxed{S^T A S} = D = [\lambda_1, \lambda_2]$$

$$S^T S = \mathbb{1}$$

$$Ax = \lambda \cancel{x} \qquad (\det(A - \lambda I) = 0)$$

$$S^T \underset{\uparrow}{A} x = \lambda S^T \cancel{x}$$

$$\mathbb{1} = S S^T$$

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

$$D \cdot \underbrace{S^T x} = \lambda \underbrace{S^T x}$$

PCA Theorem

X - multivariate random variable with zero mean $X \in \mathbb{R}^D$

d - integer $d < D$

The d-principal components of X, $y \in \mathbb{R}^D$ are defined as the d-orthogonal (uncorrelated) linear components of X

$$y_i = S_i X \in \mathbb{R}^D$$

$$S_i \in \mathbb{R}^D$$

S_i are eigenvectors of the correlation matrix.

The variance of y_i is maximum subject to maximize subject to

$$s_i^T s_i = 1 \quad / \quad s_i^T s_j = \delta_{ij}$$

$$\mathrm{var}(y_1) \geq \mathrm{var}(y_2) \geq \mathrm{var}(y_3) \ldots \mathrm{var}(y_d) > 0$$

$$s_1^* = \arg\max_{s_1 \in \mathbb{R}^D} \mathrm{var}(s_1 x)$$

subject to $s_1^T s_1 = 1$

s_i' are the orthogonal eigenvectors of $C[x]$

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