PCA & Kernel PCA

Recap

Given $x_1, \ldots, x_m \in \mathbb{R}^n$ i.i.d. random vectors

$[\,]\,[\quad]$

$$\mu = \mathbb{E}[x] \quad \text{and} \quad C = \mathbb{E}[(x-\mu)(x-\mu)^T]$$

"mean" $n \times 1$ "covariance" $n \times n$

Covariance matrix C is symmetric and semidefinite: (eigvals ≥ 0)

View 2: Find direction u_1 that captures maximizes variance of data, i.e., $\operatorname{argmax}_{u_1} \mathbb{E}[u_1^T (x-\mu)(x-\mu)^T u_1]$.

$$\Rightarrow \text{Take } u_1 \text{ s.t. } Cu_1 = \underbrace{\lambda_1}_{\text{largest eigenvalue}} u_1$$

Similarly, build up orthogonal coordinate system (ONB) $u_1, u_2, \ldots, u_n$ s.t. $Cu_k = \lambda_k u_k$ where

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0.$$

The k^{th} direction u_k maximizes the variance along u_k s.t. $u_k \perp \{u_1, \ldots, u_{k-1}\}$.

Covariance matrix C is symmetric and semidefinite: (eigvals ≥ 0)

$$C = \overset{U^T U = U U^T = I}{U \underset{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0}{\Lambda} U^T}$$

Principle Components of $x \in \mathbb{R}^n$ are entries of

$$y = U^T(x-\mu) = \begin{bmatrix} — & u_1^T & — \\ — & u_2^T & — \\ & \vdots & \\ — & u_n^T & — \end{bmatrix} \begin{bmatrix} x^{(1)} - \mu^{(1)} \\ x^{(2)} - \mu^{(2)} \\ \vdots \\ x^{(n)} - \mu^{(n)} \end{bmatrix} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}$$

<u>View 1:</u> New random vector y has

$$\mathbb{E}[y] = 0 \quad \text{and} \quad \mathbb{E}[yy^T] = \Lambda$$

<u>View 2:</u> Direction u_j maximizes $\mathrm{var}(y^{(j)})$ s.t.

$u_j \perp \{u_1, \ldots, u_{j-1}\}$, $\quad u_j = \underset{\substack{\|u_j\|=1 \\ u_j \perp \{u_1, \ldots, u_{j-1}\}}}{\operatorname{argmax}} \mathbb{E}\left[|u_j^T(x-\mu)|^2\right]$

<u>Computing Principle Components</u>

For PCA, we need μ and C. In practice, we estimate them from the data (set).

$$\tilde{\mu} = \frac{1}{m}\sum_{j=1}^{m} x_j$$

$n \ge m$

$[\ ][\ \ ]$

$$\tilde{C} = \frac{1}{m-1}\sum_{j=1}^{m}(x_j - \tilde{\mu})(x_j - \tilde{\mu})^T = \frac{1}{m-1}BB^T$$

$$B = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_m \\ | & | & & | \end{bmatrix} - \underbrace{\begin{bmatrix} | & | & & | \\ \mu & \mu & \cdots & \mu \\ | & | & & | \end{bmatrix}}_{\mu 1^T}$$

$$= X - \mu 1^T$$

Note that sample covariance C remains symm. semi-def.

<u>Approach 1:</u> (a) Collect data & subtract mean

$$B = X - \mu 1^T \qquad \overset{n\times m}{Q}\,\overset{m\times n}{R R^T Q^T}$$

(b) Then, form $\tilde{C} = \frac{1}{m-1}BB^T$ $(n\times n)$

(c) and compute $\tilde{C} = U \Lambda U^T$ $\mathcal{O}(n^3)$

<u>Approach 2:</u> (a) $B = X - \mu 1^T$

(b) Form $\tilde{D} = \frac{1}{m-1} B^T B$ $(m \times m)$

(c) and compute $\tilde{D} = V \Omega V^T$ $O(m^3)$

$$\lambda \neq 0 \qquad \overset{n \times n}{\tilde{C}} u = \lambda u \iff \overset{m \times m}{\tilde{D}} v = \lambda v$$

$$u = \frac{1}{\sqrt{m-1}} B v$$

<u>Approach 3:</u> (a) $B = X - u I$

(b) Compute $B = \overset{n \times n}{U} \overset{n \times m}{\Sigma} \overset{m \times m}{V^T}$ (Thick/Full SVD)

$$BB^T = (U \Sigma V^T)(U \Sigma V^T)^T$$
$$= U \Sigma \cancel{V^T V}^{I} \Sigma^T U^T$$
$$= U \Sigma^2 U^T \qquad \Lambda = \Sigma^2$$

$$B^T B = (U \Sigma V^T)^T (U \Sigma V^T)$$
$$= V \Sigma^T \cancel{U^T U}^{I} \Sigma V^T$$
$$= V \Sigma^2 V^T$$

In practice, compute "thin SVD" "economy"

$$\overset{n \times m}{B} = \overset{n \times m}{\tilde{U}} \overset{m \times m}{\tilde{\Sigma}} \overset{m \times m}{\tilde{V}^T} \qquad n \geq m$$

$$\overset{n \times m}{B} = \overset{n \times n}{\tilde{U}} \overset{n \times n}{\tilde{\Sigma}} \overset{n \times m}{\tilde{V}^T} \qquad n < m$$

PCA : Nonlinear Effects

To incorporate higher-order correlations, we can begin "adding" new variables. For example,

$$x_1 = \begin{pmatrix} x_1^{(1)} \\ x_2^{(2)} \end{pmatrix}, \ldots, \; x_m = \begin{pmatrix} x_1^{(1)} \\ x_2^{(2)} \end{pmatrix}$$

add $\Rightarrow$ $x_k^{(3)} = \left(x_k^{(1)}\right)^2, \; x_k^{(4)} = x_k^{(1)} x_k^{(2)}, \; x_k^{(5)} = \left(x_k^{(2)}\right)^2$

(+) The new 5×5 covariance matrix contains higher-order statistics (nonlinear effects)

(−) The "curse" of dimensionality.