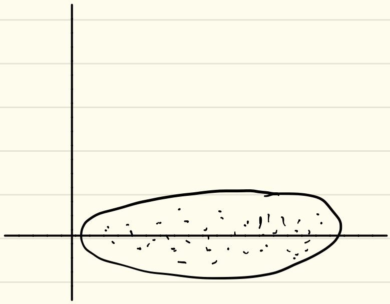
Dimensionality Reduction



$$\underline{x}_j \in \mathbb{R}^d$$
$$\downarrow \qquad , \ k < d$$
$$\underline{y}_j \in \mathbb{R}^k$$

Approx. error $= \|(x_1, x_2) - (x_1, 0)\|^2$

PCA $\rightarrow$ Principal Component Analysis

Form empirical correlation matrix $\quad \frac{1}{n} \sum_{j=1}^{n} x_j x_j^T$

Find orthonormal eigenvectors of $\hat{R} \equiv k$ largest eigenvectors of $\hat{R}$.

Project data points onto space spanned by these vectors.

$x_1, \ldots, x_n \in \mathbb{R}^d$, $\quad X_{d \times n} = [x_1 \ \ldots \ x_n]$

Let $V =$ k-dim subspace of $\mathbb{R}^d$

Let $f_1, f_2, \ldots, \underline{f}_k =$ orthonormal basis for V

$V = \text{span}(f_1, \ldots, f_k)$

$$f_i^T f_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

$$\text{Proj}_V(x_i) = \sum_{l=1}^{k} \langle x_j, f_l \rangle \, f_l$$

$$= \sum_{l=1}^{k} f_l \, (f_l^T x_j)$$

$F = [f_1 \, .. \, f_k]$

$\hookrightarrow d \times k$

$F^T F = I$

$$= \left(\sum_{l=1}^{k} f_l f_l^T \right) x_j$$

$$= F F^T x_j$$

$$\underset{\in \mathbb{R}^d}{x_j} \to \underset{\hookrightarrow k\text{-dim}}{\text{proj}_V}(x_j) = \underbrace{\underbrace{F F^T x_j}_{\in \mathbb{R}^d}}_{\hat{x}_j}$$

$\hat{x}_j$: k-dim approx to x_j

Approximation Error

$$\| x_j - \hat{x}_j \|^2 = \| x_j - \text{Proj}_V(x_j) \|^2$$

$$= \| x_j \|^2 - \| \text{proj}_V(x_j) \|^2$$

$$= \| x_j \|^2 - \| F F^T x_j \|^2 \qquad \left(\|a\|^2 = a^T a = \text{Tr}(a^T a) = \text{Tr}(a a^T) \right)$$

$$= \| x_j \|^2 - \text{Tr}\left(F^T x_j x_j^T F \right)$$

Total Approx. Error

$$= \sum_{j=1}^{n} \| x_j - \hat{x}_j \|^2$$

$$= \sum_{j=1}^{n} \| x_j \|^2 - \sum_{j=1}^{n} Tr (F^T x_j x_j^T F)$$

$$= \quad '' \quad - Tr \{ F^T (\sum_{i=1}^{n} x_j x_j^T) F \}$$

$$\text{Total error } (F) = \underbrace{(\quad)}_{\text{doesn't depend on } F} - n \, Tr \{ F^T \hat{R} F \}$$

Best k-dim subspace

$$= \underset{F : F^T F = I_k}{\operatorname{argmin}} \left[(\quad) - n \, Tr (F^T \hat{R} F) \right]$$

$$= \underset{F : F^T F = I_k}{\operatorname{argmax}} \; Tr (F^T \hat{R} F)$$

Solution $\rightarrow$ Rayleigh-Ritz Theorem

$$F = [u_1 \, .. \, u_k] \quad \text{where}$$

$$\hat{R} = U \Lambda U^T , \quad U = [u_1 \; .. \, u_k \, .. \, u_d]$$

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_d \end{pmatrix} , \quad \lambda_1 \geq \lambda_2 \geq \lambda_d \geq 0$$

Summary of PCA

Given 1) $\mathbb{X} = [x_1 \quad x_n]_{d\times n}$

2) $k \leq \operatorname{rank}(\mathbb{X})$

Steps

1) eigen decomposition of empirical correlation matrix

$$\mathbb{X}\mathbb{X}^T = U \Lambda U^T, \quad U = [u_1 \, .. \, u_d], \quad U^{-1} = U^T$$

$$\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} \qquad \lambda_k > 0 \quad (\operatorname{rank} \geq k)$$

2) Best k-dim subspace V_k^{best}

$$F = [u_1 \quad u_k]_{d\times k} \qquad \boxed{F^T F = I_k}$$

$$V_k^{best} = \operatorname{span}(f_1, \, , f_k) = \operatorname{span}(u_1, \, , u_k)$$

Finished learning prob.

3) Encode/ Embed test sample

$$x \in \mathbb{R}^d \longrightarrow y = \underbrace{F^T}_{k\times d} \underbrace{x}_{d\times 1} \in \mathbb{R}^k$$

$$= \begin{pmatrix} u_1^T x \\ \vdots \\ u_k^T x \end{pmatrix}$$

Reconstruction of approximation

$$\text{Proj}_{V_k^{best}}(x) = \hat{x} = Fy = FF^T x \in \mathbb{R}^d$$

Approx. Error

$$\| x - \hat{x} \|^2 = \|x\|^2 - \|\hat{x}\|^2 = \|x\|^2 - \|F^T x\|^2 = \|x\|^2 - \|y\|^2$$

Total approximation error of entire training set

$$= \sum_{j=1}^{n} \| x_j - \hat{x}_j \|^2 = \sum_{i=k+1}^{d} \lambda_i$$

---

eigenvector of $\mathbb{X}\mathbb{X}^T$

$$\mathbb{X}\mathbb{X}^T v = \lambda v \qquad v \neq 0 \quad \lambda > 0$$

$$(\mathbb{X}^T \mathbb{X}) \underbrace{\mathbb{X}^T v}_{u} = \lambda \underbrace{(\mathbb{X}^T v)}_{u} \qquad \text{(Kernel PCA)}$$