FoDA L21

PCA (Principal Component Analysis)

and

MDS (Multi Dimensional Scaling)

Principal Component Analysis

Input $A \in \mathbb{R}^{n\times d}$ $\quad A = \{a_1, a_2, \dots a_n\} \subset \mathbb{R}^d$

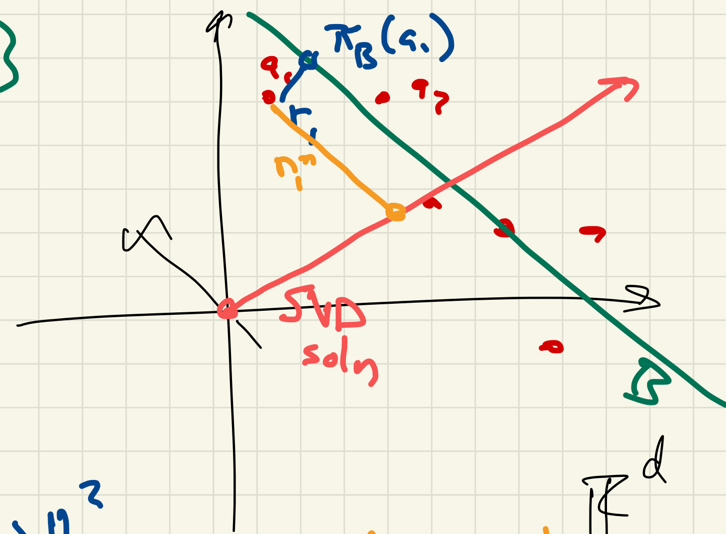
Goal B $\quad V_B = \{v_1, \dots, v_k\}$ orthogonal basis

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

$$\pi_B(x) = \sum_{j=1}^{k} v_j \langle v_j, x \rangle$$

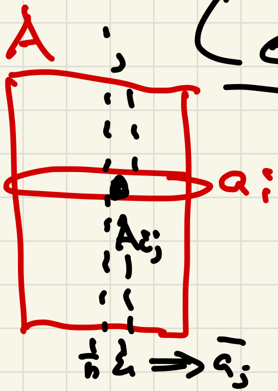
Find B minimize $\|A - \pi_B(A)\|_F^2$

$$= \sum_{i=1}^{n} \underbrace{\|a_i - \pi_B(a_i)\|}_{r_i}{}^2$$



PCA fixes by first "centering" data

SVD does not directly give soln since includes 0.



Centering the data

Shift of data so its average is the origin $\underline{0} = (0, 0, \dots 0)$

PCA = centering then SVD

$$\bar{a}_j = \frac{1}{n}\sum_{i=1}^{n} A_{ij}$$

average of jth coord. on all data.

$\bar{a} = (\bar{a}_1, \bar{a}_2, \dots \bar{a}_d) \in \mathbb{R}^d$ ← average point of all data.

$\tilde{A}$ defined so $\boxed{\tilde{A}_{ij} = A_{ij} - \bar{a}_j}$

Centering Matrix

$C_n \in \mathbb{R}^{n\times n}$

$$C_n = I_n - \frac{1}{n}\mathbb{1}\mathbb{1}^T$$

$$\begin{bmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & 1 & & \\ & 0 & & 1 & \\ & & & & \ddots \end{bmatrix} - \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \\ \frac{1}{n} & \ddots & & \\ \vdots & & & \frac{1}{n} \end{bmatrix}$$

$$= \begin{bmatrix} 1-\frac{1}{n} & & -\frac{1}{n} & \\ & 1-1/n & & \\ -\frac{1}{n} & & \ddots & \\ & & & 1-1/n \end{bmatrix}$$

$\mathbb{1} \in \mathbb{R}^{n\times 1}$

$$\mathbb{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\mathbb{1}\mathbb{1}^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & & 1 \\ \vdots & \vdots & & \vdots \end{bmatrix} \in \mathbb{R}^{n\times n}$$

$$\tilde{A} = C_n A = A - \frac{1}{n}\mathbb{1}\mathbb{1}^T A$$

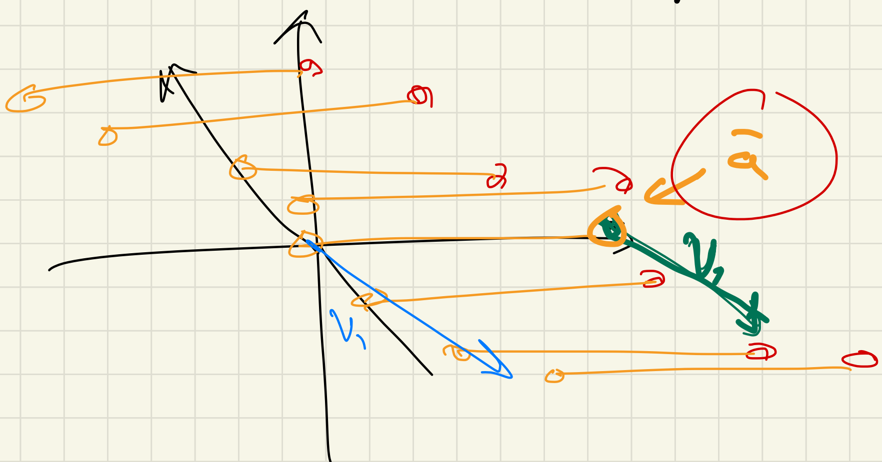
PCA

Useful for

- $k = 1, 2, 3$

1. $\tilde{A} = C_n A$
2. $U, S, V^T = \mathrm{svd}(\tilde{A})$
3. Take top k sing. values $\sigma_1 = S_{11}, \sigma_2 = S_{22}, \dots \sigma_k$

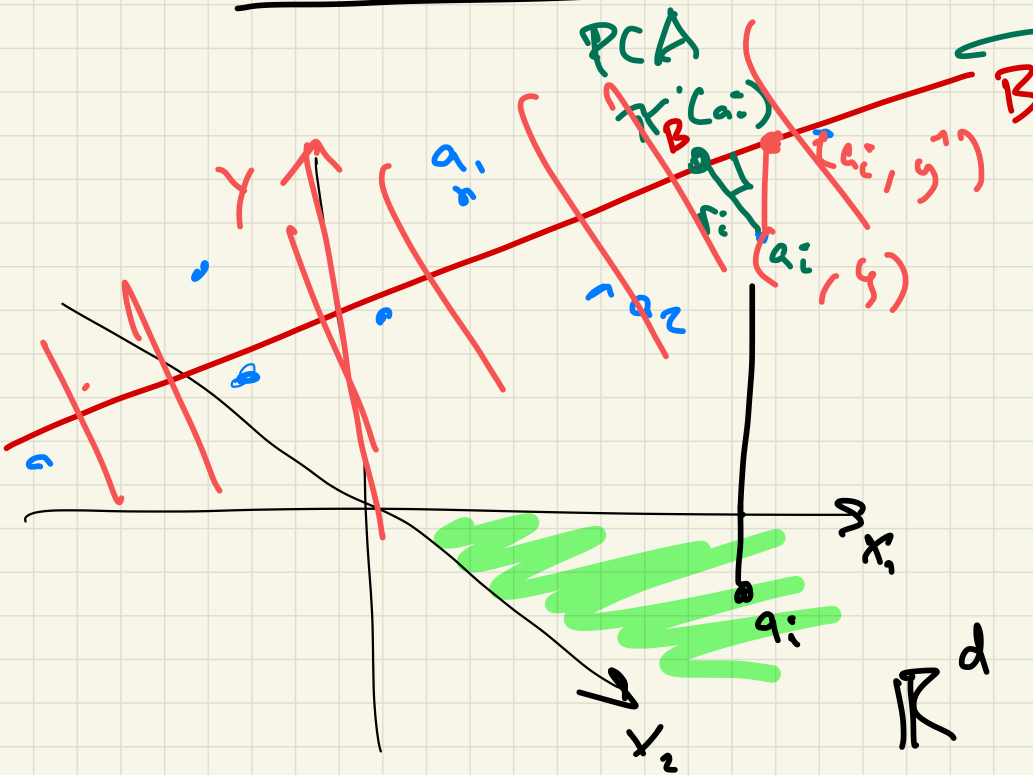
   sing vectors $\underbrace{v_1, \dots v_k}_{\text{principal components}}$



Supervised vs. Unsupervised Learning

Supervised	Unsupervised
Input (X, y)	X
Goal: Predict y using X	Find "shape" of the data (simple)
Endgoal	Subgoal
	Exploratory

Difference PCA and Regression



PCA

B rank k

$k \ll d$

Regression

$f : \mathbb{R}^d \to \mathbb{R}$

Multi Dimensional Scaling (MDS)

Input distance matrix $D \in \mathbb{R}^{n \times n}$

n objects $p_1, p_2, \ldots, p_n$ ← cities

$$D_{ij} = \text{distance}(p_i, p_j)$$

↑ cost of airplane ticket

$q_i = \mu(p_i) \in \mathbb{R}^k$

Goal Embedding μ of $p_1 \ldots p_n$ into $\mathbb{R}^k$

so $\|\mu(p_i) - \mu(p_j)\| \approx \text{distance}(p_i, p_j)$

Classical MDS

Input $D \in \mathbb{R}^{n \times n}$

Centering Matrix C_n

$$\boxed{C_n^T = C_n}$$

1. Square all distance

$$D^{(2)} \quad (D^{(2)})_{ij} = (D_{ij})^2$$

2. $M = -\frac{1}{2} C_n D^{(2)} C_n^T$

double centering

$$M \approx \tilde{A}\tilde{A}^T$$

eigenvalues

3. $[L, V] = \text{eig}(M)$

$V_k \in \mathbb{R}^{n \times k}$ $\quad L_k \in \mathbb{R}^{k \times k}$ $\begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \lambda_k \end{bmatrix}$

4. Return $Q = V_k L_k^{1/2} = \{q_1, q_2 \ldots q_n\} \in \mathbb{R}^k$

Why Classical MDS work?

Similarity Matrix $S \in \mathbb{R}^{n \times n}$

pretend there exist

$A \in \mathbb{R}^{n \times d}$

$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$

$$s_{ij} = \text{similarity}(p_i, p_j)$$

$$= \langle a_i, a_j \rangle$$

$$[A A^T]_{ij} = \langle a_i, a_j \rangle$$

$$[D^{(2)}]_{ij} = \| a_i - a_j \|^2 = \| a_i \|^2 + \| a_j \|^2 - 2 \langle a_i, a_j \rangle$$

$$\langle a_i, a_j \rangle = \frac{1}{2} \left(\| a_i \|^2 + \| a_j \|^2 - \| a_i - a_j \|^2 \right)$$

$$\langle a_i, a_j \rangle = \frac{1}{2}\left(\|a_i\|^2 + \|a_j\|^2 - \|a_i - a_j\|^2 \right)$$

Solve for $\|a_i\|^2$ and $\|a_j\|^2$

if assume $a_1 = (0, 0, 0, \ldots 0)$

$$\|a_1\|^2 = 0$$

$$\|a_i\|^2 = \|a_i - a_1\|^2 = \left[D^{(2)}\right]_{i,1}$$

$$\langle a_i, a_j \rangle = \frac{1}{2}\left(\left[D^{(2)}\right]_{i,1} + \left[D^{(2)}\right]_{j,1} - \left[D^{(2)}\right]_{i,j} \right)$$