HOMEWORK 1 : PRINCIPLE COMPONENT ANALYSIS (PCA)

Tran Thi Thu Trang

Câu 1: Tự biến đổi thuật toán PCA.

Dataset X = $\{x_1, x_2, \ldots., x_n\}$ => $x_i \in \mathbb{R}^D$

$$X = \begin{bmatrix} \_\_ x_1^T \_\_ \\ \_\_ x_2^T \_\_ \\ \vdots \\ \_\_ x_N^T \_\_ \end{bmatrix} \in \mathbb{R}^{N*D}$$

If we reduce X to M dimensions then we have :

$$B = \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_M \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{D*M}$$

$$Z = X.B = \begin{bmatrix} \_\_ x_1^T \_\_ \\ \_\_ x_2^T \_\_ \\ \vdots \\ \_\_ x_N^T \_\_ \end{bmatrix} * \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_M \\ | & | & & | \end{bmatrix} = \begin{bmatrix} x_1^T b_1 & x_1^T b_2 & & x_1^T b_M \\ x_2^T b_1 & x_2^T b_2 & & x_2^T b_M \\ \vdots & \vdots & \cdots & \vdots \\ x_N^T b_1 & x_N^T b_2 & & x_N^T b_M \end{bmatrix} = \begin{bmatrix} \_\_ z_1^T \_\_ \\ \_\_ z_2^T \_\_ \\ \vdots \\ \_\_ z_N^T \_\_ \end{bmatrix}$$

$$=> \mathbb{R}^D \to \mathbb{R}^M : \begin{cases} x_1 \to z_1 \\ x_2 \to z_2 \\ \quad \cdots \\ x_N \to z_N \end{cases}$$

Assumtion: $\boldsymbol{M_X = \frac{x_1 + x_2 + \cdots . + x_N}{N} = 0}$

$$x' = x - \mu \ \ with\ every\ data\ point$$

We have:

$$\begin{cases} x_1 \to x_1^T b_1 \\ x_2 \to x_2^T b_1 \\ \quad \vdots \\ x_N \to x_N^T b_1 \end{cases} \quad => \quad \underset{\boldsymbol{b_1}}{\boldsymbol{Max}}\ \textbf{variance} \quad (\mathbf{x_1^T} \boldsymbol{b_1}, \boldsymbol{x_2^T b_1}, \ldots \ldots, \boldsymbol{x_N^T b_1})$$

$$M_Z = \frac{\boldsymbol{x_1^T b_1} + \boldsymbol{x_2^T b_1} + \cdots + \boldsymbol{x_N^T b_1}}{\boldsymbol{N}} = \boldsymbol{b_1^T} * \frac{\sum_{i=1}^{N} \boldsymbol{x_i}}{\boldsymbol{N}} = \mathbf{0}$$

$$\boldsymbol{Var(Z)} = \frac{\sum_{i=1}^{N}\left(\boldsymbol{x}_i^T\boldsymbol{b}_1 - \boldsymbol{M}_z\right)^2}{\boldsymbol{N}} = \frac{\sum_{i=1}^{N}\left(\boldsymbol{x}_i^T\boldsymbol{b}_1\right)^2}{\boldsymbol{N}} = \frac{\sum_{i=1}^{N}\boldsymbol{b}_1^T\boldsymbol{x}_i\boldsymbol{x}_i^T\boldsymbol{b}_1}{\boldsymbol{N}} = \boldsymbol{b}_1^T\left(\frac{\sum_{i=1}^{N}\boldsymbol{x}_i\boldsymbol{x}_i^T}{\boldsymbol{N}}\right)\boldsymbol{b}_1 = \boldsymbol{b}_1^T\boldsymbol{S}\boldsymbol{b}_1$$

Where S is Covariance Matrix

We find $\max_{b_1} b_1^T S b_1$:

Constraint $\|b_1\|_2^2 = 1$

Lagrange multiplace:

$L = b_1^T S b_1 + \lambda(1 - b_1^T b_1)$

- ❖ $\frac{\partial L}{\partial b_1} = 0 \quad \Leftrightarrow \quad 2Sb_1 - 2\lambda b_1 = 0$

$$\Leftrightarrow \quad Sb_1 = \lambda b_1 \quad \rightarrow \begin{cases} b_1 : eigenvector \\ \lambda : eigenvalue \end{cases}$$

- ❖ $\frac{\partial L}{\partial \lambda} = 0 \quad \Leftrightarrow \quad 1 - b_1^T b_1 = 0$

$$\Leftrightarrow \quad b_1^T b_1 = 1$$

⇨ $Var(Z) = b_1^T S b_1 = b_1^T \lambda b_1 = \lambda$

Therefore, to maximize the variance of the low-dimensional code, we choose the basis vector associated with the largest eigenvalue principal component of the data covariance matrix.

This eigenvector is called the first principal component. The second component is the projection of data onto the eigenvector corresponding with the second largest eigenvalue.