# Homework week 1

Nguyễn Hoàng Long - DSEB 62

Ngày 27 tháng 12 năm 2022

## Example 1:

Consider a dataset $X = \begin{bmatrix} \cdot & \cdot & x_1^T & \cdot & \cdot \\ \cdot & \cdot & x_2^T & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & x_n^T & \cdot & \cdot \end{bmatrix} \in R^{N*D}$ with mean 0

We assume there exists a low dimensional compressed representation:

$$Z = XB \in R^{N*M}$$

where $B = [b_1, b_2, \dots, b_m] \in R^{D*M}$

$$\begin{aligned}
Z &= XB \\
<=> \begin{bmatrix} \cdot & \cdot & z_1^T & \cdot & \cdot \\ \cdot & \cdot & z_2^T & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & z_n^T & \cdot & \cdot \end{bmatrix} \\
&= \begin{bmatrix} \cdot & \cdot & x_1^T & \cdot & \cdot \\ \cdot & \cdot & x_2^T & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & x_n^T & \cdot & \cdot \end{bmatrix} [b_1, b_2, \dots, b_m] \\
&= \begin{bmatrix} x_1^T b_1 & x_1^T b_2 & \cdot & x_1^T b_M \\ x_2^T b_1 & x_2^T b_2 & \cdot & x_2^T b_M \\ \cdot & \cdot & \cdot & \cdot \\ x_N^T b_1 & x_N^T b_2 & \cdot & x_N^T b_M \end{bmatrix}
\end{aligned}$$

Assumption:

$$\mu_x = 0 \Longleftrightarrow E_x[x] = 0 \Longleftrightarrow XB = 0 \Longleftrightarrow E_z[Z] = 0$$

We standardize the data $x = x - \mu_x$

We need to find a matrix B that retains as much information as possible when compressing data by projecting it onto the subspace spanned by the columns $b_1, b_2, b_3, \dots, b_M$ of B.

Retaining most information after data compression is equivalent to capturing the largest amount of variance in the low-dimensional code

We start by seeking a single vector $b_1 \in R^D$ that maximizes the variance of the projected data

$$\begin{aligned}
V_1 &= \frac{1}{N} \sum_{n=1}^{N} z_{1n}^2 \\
&= \frac{1}{N} \sum_{n=1}^{N} \left(x_n^T b_1\right)^2 \\
&= \frac{1}{N} \sum_{n=1}^{N} b_1^T x_i x_i^T b_1 \\
&= b_1^T \left( \frac{\sum_{n=1}^{N} x_i x_i^T}{N} \right) b_1 \\
&= b_1^T S b_1
\end{aligned}$$

where S is the covariance matrix of X

Remark: increasing magnitude of $b_1$ increases $V_1$, we restrict all solutions to $\|b_1\|_2^2 = 1$. Then we have an optimization problem:

$$b_1^T S b_1 \Rightarrow \max$$
$$\text{subject to: } \|b_1\|_2^2 = 1$$

The Lagrangian multiplier:

$$\begin{aligned}
&L(b_1, \lambda) = b_1^T S b_1 + \lambda \left(1 - b_1^T b_1\right) \\
&\frac{\delta L}{\delta b_1} = 2b_1^T S - 2\lambda b_1^T = 0 \\
&\Longleftrightarrow b_1^T S = b_1^T \\
&\Longleftrightarrow S b_1 = \lambda b_1 \\
&\frac{\delta L}{\delta \lambda} = 1 - b_1^T b_1 = 0 \\
&\Longleftrightarrow b_1^T b_1 = 1
\end{aligned}$$

Because Sb1 = b1, we can see that b1, $\lambda$ is eigenvector and eigenvalue of S respectively.

$$V = b_1^T S b_1 = b_1^T \lambda b_1 = \lambda$$

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