1 Exercise 1:

We consider an independent and identically distributed (i.i.d.) dataset $X = \{x_1, x_2, \ldots, x_n\}$ with mean $0, x_n \in \mathrm{R}^D$. We assume there exists a low-dimensional compressed representation

$$z_n = B^T x_n \in \mathrm{R}^M$$

of x_n where projection matrix

$$B := [b_1, b_2, \ldots, b_M] \in \mathrm{R}^{D \times M}$$

$\mathbb{E}_z[z] = \mathbb{E}_x\left[B^T x\right] = B^T \mathbb{E}_x[x] = 0 \Rightarrow$ mean z is 0.
We maximize the variance of the low-dimensional code using a sequential approach. We start by seeking a single vector $b_1 \in \mathbb{R}^D$ that maximizes the variance of the projected data, i.e., we aim to maximize the variance of the first coordinate z_1 of $z \in R^M$ so that

$$V_1 = \frac{1}{N} \sum_{n=1}^{N} z_{1n}^2$$

is maximized, where

$$z_{1n} = b_1^T x_n$$

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

where S is called the data covariance matrix.
Constrained optimization problem

$$\max_{b_1} b_1^T S b_1$$
$$\text{subject to } \|b_1\|^2 = 1$$

The Lagrangian:

$$\begin{aligned} &\mathcal{L}\left(b_1, \lambda_1\right) = b_1^T S b_1 + \lambda_1 \left(1 - b_1^T b_1\right) \\ &\frac{\partial \mathcal{L}}{\partial b_1} = 2 b_1^T S - 2 \lambda_1 b_1^T, \frac{\partial \mathcal{L}}{\partial \lambda_1} = 1 - b_1^T b_1 \end{aligned}$$

Setting these partial derivaties to 0 gives us

$$\begin{aligned} b_1^T b_1 &= 1 \\ S b_1 &= \lambda_1 b_1 \end{aligned}$$

We can see that b_1, λ_1 is an eigenvector and an eigenvalue of S respectively.

$$V = b_1^T S b_1 = b_1^T \lambda_1 b_1 = \lambda_1$$

Therefore, to maximize the variance of the low-dimensional code, wechoose the basis vector associated with the largest eigenvalue principalcomponent of the data covariance matrix.