We have

$$X = \begin{bmatrix} -x_1^T- \\ -x_2^T- \\ \vdots \\ -x_n^T \end{bmatrix} \in R^{n.D}$$

Assume there is a low-dimensional compress representation, doing that by create B matrix

$$B = \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \dots & b_n \\ | & | & & | \end{bmatrix} \in R^{D*M}$$

$$\Leftrightarrow \begin{bmatrix} -z_1^T- \\ -z_2^T- \\ \vdots \\ -z_n^T \end{bmatrix} = \begin{bmatrix} -x_1^T- \\ -x_2^T- \\ \vdots \\ -x_n^T- \end{bmatrix} \cdot \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \dots & b_M \\ | & | & & | \end{bmatrix}^T$$

$$= \begin{bmatrix} x_1^T b_1 & x_1^T b_2 & \dots & x_1^T b_M \\ x_2^T b_1 & x_2^T b_2 & \dots & x_2^T b_M \\ \vdots & & & \\ x_N^T b_1 & x_N^T b_2 & \dots & x_n^T b_M \end{bmatrix}$$

We want the coordinate have the highest variance and the mean $= 0$

$$\mu_x = \frac{x_1 + x_2 + \dots x_n}{N}$$

$$x' = x - \mu_x \Rightarrow \mu_{x'} = 0$$

$$\mu_z = \frac{x_1^T b_1 + x_1^T b_2 + \dots + x_1^T b_M}{N}$$

$$= \frac{1}{N} b_1^T \sum_{i=N}^{N} x_i$$

$$= b_1^T \mu_x = 0$$

$$\mathrm{Var}_Z = \frac{1}{N}\sum_{i=N}^{N}(x_1^T b_1 - \mu_2)^2$$
$$= \frac{1}{N}\sum_{i=N}^{N}(x_1^T b_1)^2$$
$$= \frac{1}{N}\sum_{i=N}^{N} b_1^T x_i x_1^T b_i$$
$$= b_1^T\left(\frac{\sum_{i=N}^{N} x_i x_i^T}{N}\right)b_1$$
$$= \cancel{b_1^T S b_1}\; b_1^T \operatorname{cov}(X, X) b_1$$

$\max b_1^T S b_1$

subject to : $\|b_1\|^2 = 1$

Lagrangian:

$$\mathcal{L}(b_1, \lambda_1) = b_1^T S b_1 + \lambda_1(1 - b_1^T b_1)$$

$$\frac{\nabla \mathcal{L}}{\nabla b_1} = 0 \Leftrightarrow S b_1 = \lambda_1 b_1$$

$$\frac{\nabla \mathcal{L}}{\nabla \lambda_1} = 0 \Leftrightarrow b_1^T S b_1 = 1$$

$$\mathrm{Var} = b_1^T S b_1 = b_1^T \lambda b_1 = \lambda b_1^T b_1 = \lambda$$

~~We~~ 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