PAGE NO.:
DATE: / /

→ Dimensionality Reduction with PCA :-

consider an i.i.d. dataset $X = \{x_1, \dots x_N\}$, $x_n \in \mathbb{R}^D$, with mean 0 and covariance matrix

$$S = \frac{1}{N} \sum_{n=1}^{N} x_n x_n^T \quad \text{—— (1)}$$

Furthermore, we assume there exists a low-dimensional compressed representation of x_n :

$$z_n = B^T x_n \in \mathbb{R}^M \quad \rightarrow \text{Encoder.}$$

where, $B := [b_1 \dots b_m] \in \mathbb{R}^{D \times M}$
Projection matrix

And we assume that columns of B are orthonormal.
i.e, $b_i^T b_j = 0$ if and only if $i \neq j$ and $b_i^T b_i = 1$

we seek an M-dimensional subspace $U \subseteq \mathbb{R}^D$, $\dim(U) = M < D$ onto which we project the data.

projected data : $\tilde{x}_n \in U$ and their co-ordinates are z_n.

Now, ~~taking variance of compressed representation~~

$$\cancel{V_z(z)} \quad \cancel{V_z[z] = V_x[B^T(x-\mu)]}$$

$$\cancel{V_z[Z] = V_x[B^T x - B^T \mu]}$$
$$\cancel{= V_x[B^T}$$

Now, ~~taking~~ Maximizing variance of the first coordinate z_1 of $z \in R^M$

$$V_1 := V_1[z_1] = \frac{1}{N}\sum_{n=1}^{N} z_{1n}^2 \quad , \quad z_{1n} = b_1^T x_n$$

$$V_1 = \frac{1}{N}\sum_{n=1}^{N} (b_1^T x_n)^2$$

$$= \frac{1}{N}\sum_{n=1}^{N} b_1^T x_n x_n^T b_1$$

$$= b_1^T \underbrace{\left(\frac{1}{N}\sum_{n=1}^{N} x_n x_n^T\right)}_{S} b_1$$

$$V_1 = b_1^T S b_1 \quad \text{—— (2)}$$

Maximum variance can be found by constrained optimization problem

$$\max_{b_1} b_1^T S b_1$$

Subject to $||b_1||^2 = 1$

The Lagrangian is obtained as:

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

DATE: / /

Take partial derivatives of $\mathcal{L}$ w.r.t. b_1 and λ_1

$$\frac{\partial \mathcal{L}}{\partial b_1} = 2b_1^T S - 2\lambda_1 b_1^T = 0^T$$

$$S b_1 = \lambda_1 b_1 \quad \text{——— } (3)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = 1 - b_1^T b_1 = 0 \Rightarrow b_1^T b_1 = 1 \text{ — } (4)$$

put value of Sb_1 in eq^n (2)

$$V_1 = b_1^T S b_1 = \lambda_1 b_1^T b_1$$

$$V_1 = \lambda_1 \quad \text{——— } (5)$$

To maximize the variance of the low-dimensional data, we choose the basis vector associated with the largest eigen value of the covariance matrix of data.

We can reconstruct the data points by,

$$\tilde{x}_n = b_1 z_{1n} \quad \leftarrow \text{Decoder}$$

$$= b_1 b_1^T x_n \in R^D$$