ML2 dseb62 w1

Nguyen Tuan Duy - 11204971 - dseb62

December 2022

1 Problem 1

Consider a dataset $X = \begin{bmatrix} . & . & x_1^T & . & . \\ . & . & x_2^T & . & . \\ . & . & . & . & . \\ . & . & x_n^T & . & . \end{bmatrix} \in R^{N*D}$ with mean 0, we assume there exists a low dimensional compressed representation:

$$Z = XB \in R^{N*M} \tag{1}$$

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

$$\begin{aligned} Z &= XB & (2) \\ &= \begin{bmatrix} . & . & x_1^T & . & . \\ . & . & x_2^T & . & . \\ . & . & . & . & . \\ . & . & x_n^T & . & . \end{bmatrix} [b_1, b_2, ..., b_m] & (3) \\ &= \begin{bmatrix} x_1^T b_1 & x_1^T b_2 & . & x_1^T b_M \\ x_2^T b_1 & x_2^T b_2 & . & x_2^T b_M \\ . & . & . & . \\ x_N^T b_1 & x_N^T b_2 & . & x_N^T b_M \end{bmatrix} & (4) \end{aligned}$$

We assumed that:

$$\mu_x = 0 \iff E_x[x] = 0 \iff XB = 0 \iff E_z[Z] = 0 \tag{5}$$

To satisfy this assumption, we standardize the data $x = x - \mu_x$
Our goal is to find matrix B that retains as much information as possible when compressing data by projecting it onto the subspace spanned by columns $b_1, b_2, b_3, ..., b_M$ of B. This problem 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

$$V_1 = \frac{1}{N}\sum_{n=1}^{N} z_{1n}^2 \tag{6}$$

$$= \frac{1}{N}\sum_{n=1}^{N} (x_n^T b_1)^2 \tag{7}$$

$$= \frac{1}{N}\sum_{n=1}^{N} b_1^T x_i x_i^T b_1 \tag{8}$$

$$= b_1^T (\frac{\sum_{n=1}^{N} x_i x_i^T}{N}) b_1 \tag{9}$$

$$= b_1^T S b_1 \tag{10}$$

where S is the covariance matrix of X
Because increasing magnitude of b_1 increases V_1, we restrict all solutions to $||b_1||_2^2 = 1$. Then we have an optimization problem:

$$b_1^T S b_1 \Rightarrow max \tag{11}$$

$$\text{subject to: } ||b_1||_2^2 = 1$$

The Lagrangian multiplier:

$$L(b_1, \lambda) = b_1^T S b_1 + \lambda(1 - b_1^T b_1) \tag{12}$$

$$\frac{\delta L}{\delta b_1} = 2b_1^T S - 2\lambda b_1^T = 0 \tag{13}$$

$$\iff b_1^T S = b_1^T \tag{14}$$

$$\iff Sb_1 = \lambda b_1 \tag{15}$$

$$\frac{\delta L}{\delta \lambda} = 1 - b_1^T b_1 = 0 \tag{16}$$

$$\iff b_1^T b_1 = 1 \tag{17}$$

based on (15), we can see that b1, λ is eigenvector and eigenvalue of S

$$V = b_1^T S b_1 = b_1^T \lambda b_1 = \lambda \tag{18}$$

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