Machine Learning 2 - HW1 - PCA

Chu Duc Trung - 11207309

Ngày 27 tháng 12 năm 2022

1 Problem 1

We consider an independent and identically distributed dataset X $= \{x_1, x_2, \ldots, x_n\}$ with mean $0, x_n \in \mathrm{R}^D$. The goal of PCA is to map this space to a space which has dimensionality of $M(M < D)$

1) Mean subtraction and standardization, get the data to the normaldistribution.

By subtracting the respective means from the numbers in the respective column to produces a dataset whose mean is zero.

$$Let : \mu = \frac{1}{N} \sum_{i=1}^{N} x^{(i)}$$

If $\mu_x \neq 0$ we need to replace each x$^{(i)}$ with $x^{(i)} - \mu$

2) Eigendecomposition of the covariance matrix: Compute the data covariance matrix and its eigenvalues and corresponding eigenvectors. Because covariance matrix is symmetric eigenvectors are orthogonal.

We start by seeking a single vector $b1 \in R^D$ that maximizes the variance of the projected data. So we have projections of x_i on b_1

$$\mathrm{Proj}_{b_1}(x_i) = b_1^T x_i b$$

and Mean of projections $= b_1^T \bar{x} b$ where b is unit vector which has magnitude equals 1 .

Then we have variance of projections

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

where S is covariance matrix

3) Projection: Project the data onto the eigenvectors.

Max: $b_1^T S b_1$

s.t: $b_1^T b_1 = 1$

Using lagrange multiplier, we have a new objective function

$$L\left(b_1, \lambda\right) = b_1^T S b_1 - \lambda\left(1 - b_1^T b_1\right)$$

$$\frac{\partial L}{\partial b_1} = 2Sb_1 - 2\lambda b_1 = 0 \Leftrightarrow Sb_1 = \lambda b_1 (*)$$

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

$$\frac{\partial L}{\partial \lambda} = 1 - b_1^T b_1 = 0 \Leftrightarrow b_1^T b_1 = 1$$

Hit both sides of $(*)$ by b_1^T, so we get

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

Therefore, 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.