Machine Learning 2: Principle Component Analysis (PCA)

Tran Kha Uyen

Ngày 27 tháng 12 năm 2022

Câu 1: Tự biến đổi lại toán thuật toán PCA

Trả lời:
Given $\{x_n\}$ where $n = 1, 2, .., N$ with dim D. The goal of PCA is to map this space to a space which has dimensionality of M $(M < D)$

1. Normalize the data:
This is done by subtracting the respective means from the numbers in the respective column. This produces a dataset whose mean is zero.

$$\mu_x = 0 = \frac{x_1 + x_2 + ... + x_n}{N}$$

If $\mu_x \neq 0$ we need to normalize the data

$$x' = x - \mu_x$$

2. Projections of x:
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

$$Proj_{b_1}(x_i) = b_1^T x_i b$$

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

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N}(b_1^T x_n - b_1^T \bar{x})^2 &= \frac{1}{N}\sum_{n=1}^{N}[b_1^T(x_n - \bar{x})] && (1)\\
&= b_1^T[\frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x})(x_n - \bar{x})^T]b_1 && (2)\\
&= b_1^T S b_1 && (3)\\
& && (4)
\end{aligned}$$

where S is covariance matrix

3. Maximize variance of projections:

$$\text{Max: } b_1^T S b_1$$
$$\text{s.t: } b_1^T b_1 = 1$$

Using lagrange multiplier, we have a new objective function

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

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

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

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

$\Rightarrow$ To maximize the left side (the initial problem to maximize), we need to maximize the right side as well.
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.
This eigenvector is called the first principal component. The second component is the projection of data onto the eigenvector corresponding with the second-largest eigenvalue.