Derivation of PCA: maximum variance formulation

Data: $\{x_i\}$, $i = 1, 2, \cdots N$, $x_i \in \mathbb{R}^D$

mean: $\bar{x} = \frac{1}{N} \sum_i x_i$

$\boxed{\text{Assume:}}$ we shift data such that the mean of the new data is zero, namely

$$x_i \leftarrow x_i - \bar{x}$$

Define: covariance matrix

$$S = \frac{1}{N} \sum_{i=1}^{N} x_i x_i^T \in \mathbb{R}^{D \times D}$$

design matrix

$$X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{pmatrix} \in \mathbb{R}^{N \times D}$$

$$\Rightarrow \quad S = \frac{1}{N} X^T X$$

We want to find a projection u_1 such that the projected variance is maximized.

Projection of x_i is $u_1^T x_i$

the mean of the projection is

assumption

$$\frac{1}{N} \sum_i u_1^T x_i = u_1^T \sum_i \frac{1}{N} x_i = 0$$

Thus, the variance is

$$\frac{1}{N} \sum_i (u_1^T x_i)^2 = \frac{1}{N} \sum_i u_1^T x_i x_i^T u_1$$

$$= u_1^T \frac{1}{N} \sum_i x_i x_i^T u_1 = u_1^T S u_1$$

Namely, we want to

$$\max_{u_1} \quad u_1^T S u_1$$

This is not well-defined, as

$$u_1^T S u_1 \geq 0 \quad (\text{since } S \text{ is positive semi-definite})$$

Thus,

$$(2u_1^T) S (2u_1) \geq u_1^T S u_1$$

Namely, u_1 can be scaled such that

$u_1^T S u_1$ is arbitrarily large.

Thus, we restrict u_1 such that $\|u_1\|_2^2 = 1$

In other words, We solve u_1 in the following

$$\boxed{\begin{array}{ll} \max\limits_{u_1} & u_1^T S u_1 \\ \text{such that} & \|u_1\|_2^2 = 1 \end{array}}$$

First Principal Component

Define:

$$\begin{aligned} L &= u_1^T S u_1 + \lambda(1 - \|u_1\|_2^2) \\ &= u_1^T S u_1 + \lambda(1 - u_1^T u_1) \end{aligned}$$

$$\Rightarrow \frac{\partial L}{\partial u_1} = 2Su_1 - 2\lambda u_1 = 0$$

$$\Rightarrow \boxed{Su_1 = \lambda u_1}$$

Namely, (λ, u_1) is the (eigenvalue, eigenvector) of S !

But $S \in \mathbb{R}^{D \times D}$, so it has D

(eigenvalue, eigenvector) pairs.

$\boxed{\text{Which to choose?}}$

The objective function is

$$\max u_1^T S u_1 = u_1^T (\lambda u_1) = \lambda (u_1^T u_1) = \lambda$$

$\boxed{\text{Thus, we shall choose the eigenvalue } \lambda_1 \text{ that is maximal!}}$

$\boxed{\text{Other Principal Components}}$

Find u_2, such that

$u_2^T u_1 = 0$ ← orthogonal to u_1

and $u_2^T S u_2$ is maximized.

This results in

$$S u_2 = \lambda u_2$$

Such that $u_2^T u_1 = 0$

Namely, (λ, u_2) should be another (eigenvalue, eigenvector) pair different from (λ_1, u_1)

$\boxed{\text{Easy to see, the 2nd principal component should be the second-largest eigenvalue } \lambda_2 \text{ and its eigenvector of } S}$