# The Principle of PCA

WiZ

5th June 2019

## 1 Problem

We want data $x$ map to hyperplane $u$, which contains by vector $u_1$ and $u_2$, the problem is: **how to minimal the distance from $x$ to $u_1$.**

## 2 Transform

We transform the minimal distance problem to the maximal projection problem, others said maximal variance, that is:

$$max(|\vec{x}_i \cdot \vec{u}_1|) \tag{1}$$

## 3 Derivation

Assume data$\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\}$ after centralizing. The problem is to maximal:

$$\frac{1}{n}\sum_{i=1}^{n}|\vec{x}_i \cdot \vec{u}_1| \tag{2}$$

The problem equals to maximal:

$$\frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i \cdot \vec{u}_1)^2 \tag{3}$$

In this way, we have:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}|\vec{x}_i\cdot\vec{u}_1|^2 &= \frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i\cdot\vec{u}_1)^2 && (4)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(x_i^T u_1\right)^2 && (5)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(x_i^T u_1\right)^T\left(x_i^T u_1\right) && (6)\\
&= \frac{1}{n}\sum_{i=1}^{n}u_1^T x_i x_i^T u_1 && (7)\\
&= \frac{1}{n}u_1^T\left(\sum_{i=1}^{n}x_i x_i^T\right)u_1 && (8)\\
&= \frac{1}{n}u_1^T X X^T u_1 && (9)
\end{aligned}$$

(4)→(5):$\vec{x}_i\cdot\vec{u}_1 = x_i^T u_1$
(5)→(6):$|AA^T| = |A||A^T| = |A||A| = |A|^2$
(8)→(9):$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$, $X^T = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}$ and $XX^T = \sum_{i=1}^{n} x_i x_i^T$

To solve the maximal $\frac{1}{n}u_1^T XX^T u_1$, we first prove $XX^T$ is a positive semi-definite matrix.(equal to prove all eigenvalue are greater than 0)

*Proof.* $XX^T$ is a positive semi-definite matrix.
Assume: $XX^T$ have one eigenvalue $\lambda$,corresponding eigenvector $\xi$,then:
$\Rightarrow \quad XX^T\xi = \lambda\xi$
$\Rightarrow \quad \left(XX^T\xi\right)^T\xi = (\lambda\xi)^T\xi$
$\Rightarrow \quad \xi^T XX^T\xi = \lambda\xi^T\xi$
$\Rightarrow \quad \xi^T XX^T\xi = \left(X^T\xi\right)^T\left(X^T\xi\right) = \left\|X^T\xi\right\|^2 = \lambda\xi^T\xi = \lambda\|\xi\|^2$
$\Rightarrow \quad \left\|X^T\xi\right\|^2 = \lambda\|\xi\|^2 \to \lambda \geq 0$

□

# 4 Solution

After provement, $\frac{1}{n}u_1^T XX^T u_1$ has maximal value. how to solve it?

## 4.1 Method 1: Lagrange multiplier

Objective function and constraint conditions constitute the maximum problem:

$$\left\{\begin{array}{c} \max\left\{u_1^T XX^T u_1\right\} \\ u_1^T u_1 = 1 \end{array}\right. \tag{10}$$

Construct Lagrange function:

$$f(u_1) = u_1^T X X^T u_1 + \lambda \left(1 - u_1^T u_1\right)$$

Take the derivative of u1 and set it equal to 0:

$$\frac{\partial f}{\partial u_1} = 2XX^T u_1 - 2\lambda u_1 = 0 \rightarrow XX^T u_1 = \lambda u_1$$

Obviously, u1 is the eigenvector, take it to formor equation:

$$u_1^T X X^T u_1 = \lambda u_1^T u_1 = \lambda$$

So, if you take the maximum eigenvalue, you get the maximum target value. You might wonder: why is the first derivative zero, the maximum? Find the second derivative:

$$\frac{\partial^2 f}{\partial u_1} = 2\left(XX^T - \lambda I\right)$$

when $\lambda$ taking to the maximal one, $XX^T - \lambda I$ is a negative semidefinite matrix.

Therefore, the objective function maximizes on the eigenvector corresponding to the maximum eigenvalue. The first principal axis is the direction of the eigenvector corresponding to the first eigenvalue, the second principal axis is the direction of the eigenvector corresponding to the second largest eigenvalue, and so on.

### 4.2 Method 2: Singular value decompositionr

For a vector, the square of its two norm (modulus length) is:

$$\|x\|_2^2 = < x, x > = x^T x$$

The objective function is converted to:

$$u_1^T X X^T u_1 = \left(X^T u_1\right)^T \left(X^T u_1\right) = < X^T u_1, X^T u_1 > = \left\|X^T u_1\right\|_2^2$$

The problem becomes, for a matrix, which transforms a vector, how can we make the modulus scale of the transformed vector the largest (two norm)?

Introduce a theorm:

$$\frac{\|Ax\|}{\|x\|} \leq \sigma_1(A) = \|A\|_2$$

$\sigma_1(A)$ denotes the maximum singular value of matrix A, the singular value of a matrix A is AA (or A A) eigenvalue squared (previously proved eigenvalue¿=0).

We first make a provement:

*Proof.*

$$\frac{\|Ax\|}{\|x\|} \leq \sigma_1(A) = \|A\|_2$$

Assume: Symmetrical matrix$A^T A \in \mathbb{C}^{n \times n}$ have n eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, corresponding eigenvector$\xi_1, \xi_2, \cdots, \xi_n$, take one vector $x$:

$$x = \sum_{i=1}^{n} \alpha_i \xi_i$$

And we have:

$$\|x\|_2^2 =< x, x >= \alpha_1^2 + \cdots + \alpha_n^2$$

Also:

$$\|Ax\|_2^2 =< Ax, Ax >= (Ax)^T Ax = x^T A^T Ax =< x, A^T Ax >$$

Take $x = \sum_{i=1}^n \alpha_i \xi_i$ to above equation:

$$\begin{aligned} < x, A^T Ax > &=< \alpha_1\xi_1 + \cdots + \alpha_n\xi_n, \alpha_1 A^T A\xi_1 + \cdots + \alpha_n A^T A\xi_n > \\ &=< \alpha_1\xi_1 + \cdots + \alpha_n\xi_n, \lambda_1\alpha_1\xi_1 + \cdots + \lambda_n\alpha_n\xi_n > \\ &=< \lambda_1\alpha_1^2 + \cdots + \lambda_n\alpha_n^2 > \\ &\le \lambda_1 \left(\alpha_1^2 + \cdots + \alpha_n^2\right) = \lambda_1\|x\|_2^2 \end{aligned}$$

Therefore:

$$\frac{\|Ax\|_2}{\|x\|_2} \le \sqrt{\lambda_1} = \sigma_1$$

Notes: Since the unit eigenvectors are orthogonal, the same inner product is 1, and the different one is 0. □

Clearly, when $x = \xi_1$, $A$ have a maximum singular value,

$$\|A\xi_1\|_2^2 =< \xi_1, A^T A\xi_1 >=< \xi_1, \lambda_1\xi_1 >= \lambda_1$$

$$\|A\xi_1\|_2 = \sqrt{\lambda_1} = \sigma_1$$

Back to our proposal:

$$u_1^T XX^T u_1 = \left\|X^T u_1\right\|_2^2$$

Replace $A$ with $X^T$, $u_1$is our maximum eigenvalues' eigenvector, the second one and so on.

## 5 Conclution

Firstly, problem transform.
Next, some basic knowledge about Matrix.
Last, more patient and careful. [1]

[1] Have fun in study