Abstract

In the last issue, we learned the derivation process of the formula of PCA, but it is more troublesome in practical use. The co-variance matrix needs to be obtained first, and then apply singular value decomposition to it. Therefore, the more common method is to directly perform singular value decomposition on the centralized data set.

In addition, using principal component analysis, we finally get a new coordinate base. To reduce the dimension of the data set, we also need to project the coordinates. Therefore, this issue will introduce a similar but simpler method: principal coordinate analysis ($PCoA$)

Algorithm

A. SVD && PCA

In the previous issue, we derived a simplified form of the co-variance matrix:

$$S = \frac{1}{N} X^T H X \tag{10}$$

At the same time, we also derived that the central matrix $H^2 and H^T$ are their own H.

Therefore:

$$S = \frac{1}{N} X^T H^T H X \tag{11}$$

Because we can perform singular value decomposition on any matrix, we have:

$$HX = U\Sigma V^T \tag{12}$$

Therefore, it is substituted into the covariance matrix:

$$S = V\Sigma U^T U\Sigma V^T \tag{13}$$

As we all know:

$$U^T U = I \quad V^T V = VV^T = I \quad \Sigma diagonal matrix : \tag{14}$$

Therefore:

$$S = V\Sigma^2 V^T \tag{15}$$

Here, we find that we can get the eigen value $sigma$ and eigen vector V of the co-variance matrix by singular value decomposition of the centralized data set.

We calculate HXV to get the projected coordinates.

B.$PCoA$

Let's reverse the form of s and construct a matrix:

$$T = HXX^TH^T \tag{16}$$

Similar to the above process, we get:

$$\begin{aligned} T &= HXX^TH^T \\ &= U\Sigma V^TV\Sigma U^T \\ &= U\Sigma^2 U^T \end{aligned} \tag{17}$$

We will slightly derive the projected coordinates:

$$HXV = U\Sigma V^TV = U\Sigma \tag{18}$$

Therefore, the principal coordinate analysis can directly calculate the projection coordinates