Abstract

在上期，我们学习了 PCA 公式的推导过程，但该公式在实际使用中较为麻烦。需要先求出协方差矩阵，再对其进行奇异值分解。因此更常用的方法是直接对中心化的数据集进行奇异值分解。

此外，使用主成分分析，我们最终得到的是新的坐标基，要对数据集进行降维，还需要再进行坐标的投影。因此本期将介绍一种相似但更为简便的方法:主坐标分析 ($PCoA$)

Algorithm

A.SVD && PCA

在上一期中，我们推导出了协方差矩阵的简化形式：

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

同时，我们也顺带推导出中心矩阵 H^2以及H^T 都是其本身 H。

因此得到：

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

又因为我们可以对任何矩阵进行奇异值分解，因此我们有：

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

因此，代入协方差矩阵中：

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

我们知道：

$$U^T U = I \quad V^T V = V V^T = I \tag{14}$$

Σ 为对角矩阵

因此：

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

写到这里，我们发现，只需对中心化的数据集进行奇异值分解，我们就可以得到协方差矩阵的特征值 Σ 和特征向量 V。

我们计算 HXV 即可得到投影后的坐标。

B.$PCoA$

下面我们对 S 的形式做一下颠倒，构造一个矩阵:

$$T = H X X^T H^T \tag{16}$$

与上述过程相似，我们得到:

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

我们将投影后的坐标稍加推导：

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

因此主坐标分析可以直接求出投影坐标