

Linear Dimension Reduction

2022-08-10

线性降维方法主要就是 *PCA*(*principal component analysis*) 以及 *SVD*(*singular value decomposition*), 以及 *PCA* 的扩展 *MDS*(*Multidimensional Scaling*)。 *PCA* 与 *MDS* 的差别在于 *PCA* 考虑的是样本的特征, 寻求在低维空间下保留方差较大的特征的信息, 所以通过对特征之间的协方差矩阵进行特征值分解矩阵, 得到在低维空间下能够保留方差较大特征的正交基, 是对特征的线性组合, 而 *MDS* 考虑的是样本之间的相似度矩阵, 通过对相似度矩阵进行特征分解找到在低维空间下能够保留样本最大距离的正交基。

1 PCA

样本矩阵 $m * n$, n 个样本, m 个特征。

1. 计算样本矩阵协方差矩阵。
2. 协方差矩阵特征分解, 找到使得协方差最大的方向。
3. 特征向量单位化 (使得其长度为 1), 特征向量与样本做点积, 因为特征向量长度为 1, 其点积就是样本在特征向量方向上的投影长度, 代表样本在这个方向上的坐标, 选择多少个特征向量就为降到多少维。

2 MDS

MDS 是一类基于样本距离进行映射的方法。*MDS* 针对不同类型的数据有不同的方法。*PCoA* 是 *MDS* 针对于数值数据分析的方法。

2.1 PCoA

考虑一个样本的 *dissimilarity matrix*, 也就是一个含有样本之间距离或不相似度量的矩阵, 记为 D 。

- *The Torgerson method*

首先对 D 进行 *double centering* 得到 *double-centered matrix*, 记为 B 。然后对矩阵 B 进行奇异值分解。

double centering1):

$$D^2 = [d_{ij}^2]$$

double centering2):

$$B = -\frac{1}{2}CD^2C$$

其中 $C = I - \frac{1}{n}J_n$

- *The iterative method*

该方法更加实用, 可以用于非欧式距离矩阵。

$$Stress_D(x_1, x_2, \dots, x_N) = \left(\sum_{i \neq j \neq 1, \dots, N} (d_{ij} - \|x_i - x_j\|)^2 \right)^{\frac{1}{2}}$$

(APPENDIX) Appendix

3 矩阵分析

在进行讨论之前, 需要明白的是: 1) 向量的每个数值代表了在每个基下的分量。2) 当我们从不同视角观察 (不同的基), 对同一个物 (向量) 进行观察, 看到的是不一样的 (向量的表达不一样), 尽管在本质上它们是同一个物 (向量)。3) 矩阵相乘的形式是统一的, 其结果也是确定的, 但在不同的问题之下, 可以有不同的理解,

3.1 空间变换

考虑一个向量 a , $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$, 左乘一个矩阵 M , $\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$

首先从空间变化的角度理解矩阵相乘。所谓的空间变化, 就是对原空间的映射, 即原空间 A 经过变化 M , 变为 B 。这里假设的是, 无论是这个向量, 还是矩阵都是从同一组基的视角来表示的, 我们将这组基记为 $[i \ j]$, 向量 $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ 在基 i 上的分量就是 -1 , 在基 j 上的分量就是 2 。这组基下的所有向量都是依赖于该基的, 所以空间的变换直接用基的变换表示即可。矩阵 M 的列向量表示的含义就是经过空间变化以后, 原来的一组基向量在基 i, j 下的表示。因为变化是整个空间的变化, 所有向量的变化都是一致的, 所以我们只需要用原向量在原基的每个分量乘以新的基即可:

$$-1 * \begin{bmatrix} 3 \\ 1 \end{bmatrix} + 2 * \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

结论:

1. 对于左乘的矩阵 A 的每一列, 我们可以认为是原基在空间变换 A 以后用原基表示的原基的映射。

- 行列式

行列式 $\det(A)$ 给出了由 A 表示的映射引起的缩放因子和方向。数值代表缩放因子, 正负号代表变换的方向。当行列式等于 1 时, 由 A 定义的线性变换是等值的和保持方向的。

如果一个 n 阶方阵的行列式为 0 , 也就证明这个行列式不满秩。也就是说组成行列式的列向量是线性相关的, 那么这 n 个列向量张成的空间便是 n 维空间的一个投影, 维度等于秩。

$\det(A) = 0$, 变换 A 使得空间的维度被压缩了。

- 非方阵

非方阵没有行列式。按照空间变化, 非方阵一定使得空间的维度发生了变化, 无法衡量变换的大小, 因此没有行列式。例如一个 3×2 的矩阵可以将一个二维平面的向量映射为三维空间中的一个平面上的向量。而一个 2×3 的矩阵可以将一个三维空间的向量映射为二维平面上的一个向量, 考虑整个三维空间就是将原空间做了一个投影。

- 矩阵的逆

矩阵不一定存在逆。可以求逆的矩阵叫做可逆矩阵，也叫非奇异矩阵。矩阵为非奇异矩阵的充要条件是矩阵存在行列式且不为 0。可以这样理解：矩阵的作用是空间变换，其实就类似于一个函数。（这里说映射更标准些）矩阵求逆就类似于求反函数。当矩阵为不存在行列式或者行列式为 0 时，代表变换发生了降维，其映射关系不是一对一的关系，是多对一的关系，因此无法求逆或没有意义。

A^{-1} 所代表的变换恰好是 A 变换的逆过程。

求解 $Ax = b$ 的几何意义，就是找到一个向量 x 使得在 A 的变换下， x 被映射为 b 。如果 A 为满秩矩阵，则有唯一解 $x = A^{-1}B$ ，也就是对 B 施加逆变换即可找到 x

3.2 基变换

矩阵相乘同样可以从基变化的角度看。假设一个向量 a ，其在基 u, v 下的表示为 $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ 。如果我们想从另一组基 i, j 来表示应该怎么做呢？实际上我们只需要把基 u, v 用基 i, j 表示即可。假设基 u 用 i, j 表示为 $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ ， v 用 i, j 表示为 $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ ，构成矩阵 P 。那么把向量变化到 i, j 的视角下，就是简单的矩阵左乘向量（这里用到了空间变化的思想，即左乘矩阵的列是基向量）：

$$2 * \begin{bmatrix} -1 \\ 2 \end{bmatrix} + 1 * \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

如果我们想把基 i, j 下的向量从基 u, v 的视角来观察呢？答案是对于 i, j 视角下的向量我们只需要左乘 P^{-1} ，即逆矩阵可以理解为两种视角的相互转换。

既然从不同的视角看向量有不同的表示，从不同的视角看空间变化，也有不同的表示。考虑在基 u, v 下的一个空间变化表示为 M ，那么在基 i, j 下，该变化怎么表示呢。我们可以考虑先变换到基 i, j 的视角，然后进行变换，最后回到 u, v 的视角。即：

$$M = P^{-1}AP \rightarrow A = PMP^{-1}$$

结论：

1. 对于左乘矩阵 A ，我们可以认为其每一列是用新基视角表示原基的表示。

2. 这说明对于形如 $A = PMP^{-1}$ 的式子，我们可以直观的认为是视角转换下的同一变换。

符号约定：小写字母 x, i, j, u, v 代表列向量。

基变化，如何把基 (i, j) 下的坐标转化为另一组基 (u, v) 下的坐标，实际上只需要将基 (i, j) 用基 (u, v) 进行表示，然后左乘向量 x 即可。