Dimension Reduction

Principal Components Anaylsis

Maximum Variance Perspective

Lemma:
样本方差矩阵：

$$S=\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})(x_i-\overline{x})^T=\frac{1}{N}X^THX$$

其中

$$H_N=I_N-\frac{1}{N}1_N1_N^T$$

PCA的核心思想是对原始特征空间的重构，把一组线性相关的变量通过正交变换变换成线性无关的变量，以获得最大投影方差和最小重构距离。
为了使得我们的数据均值为0，方便计算，首先我们要对数据进行中心化处理：

$$x_i:=x_i-\overline{x}$$

假设我们的投影方向是$\vec{u_1}\left(|\vec{u_1}|=1\right)$，在这方向上我们有最大投影方差。我们的数据向量$\vec{x_i}-\overline{x}$在$\vec{u_1}$上的投影可以写作：

$$|\vec{x_i}-\overline{x}|\cos\theta=(\vec{x_i}-\overline{x})\cdot\vec{u_1}=(\vec{x_i}-\overline{x})^Tu_1$$

因此投影方差为：

$$
\begin{aligned}
J(u_1) &= \frac{1}{N}\sum_{i=1}^{N}((\vec{x}_i-\overline{x})^T u_1)^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} u_1^T(\vec{x}_i-\overline{x})(\vec{x}_i-\overline{x})^T u_1 \\
&= \frac{1}{N}u_1^T\sum_{i=1}^{N}(\vec{x}_i-\overline{x})(\vec{x}_i-\overline{x})^T u_1 \\
&= u_1^T S u_1
\end{aligned}
$$

于是我们的问题便转换为：

$$
\begin{cases}
\hat{u_1} = \arg\max\limits_{u_1} u_1^T S u_1 \\
u_1^T u_1 = 1
\end{cases}
$$

这个优化问题可以用拉格朗日乘数法求解：

$$
\begin{aligned}
&\mathcal{L}(u_1,\lambda) = u_1^T S u_1 + \lambda(1-u_1^T u_1) \\
&\frac{\partial \mathcal{L}}{\partial u_1} = 2Su_1 - 2\lambda u_1 = 0
\end{aligned}
$$

不难看出λ应为方差矩阵S的特征值，u_1为对应的特征向量。如果需要降到q维可以取前q个特征值和特征向量。

Minimum Error Perspective

下面从最小重构代价方面考虑PCA，首先依然要对数据进行中心化：

$$
x_i := x_i - \overline{x}
$$

我们假设原来的向量可以由基底$u_1, u_2, ..., u_p$表示，即：

$$
x_i = \sum_{i=1}^{p}(x_i^T \cdot u_i)u_i
$$

如果我们要降到q维，那么降维后数据可以表示为：

$$\hat{x_i} = \sum_{i=1}^{q}(x_i^T \cdot u_i)u_i$$

那么，我们的最小重构代价皆可以表示成每个数据向量之差的模的平均值：

$$\begin{aligned}
J &= \frac{1}{N}\sum_{i=1}^{N} \parallel x_i - \hat{x_i} \parallel^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} \parallel \sum_{j=q+1}^{p}(x_i^T u_j)u_j \parallel^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=q+1}^{p}(x_i^T u_j)^2 \\
&\triangleq \frac{1}{N}\sum_{i=1}^{N}\sum_{j=q+1}^{p}((x_i - \hat{x_i})^T u_j)^2 \\
&= \sum_{j=q+1}^{p}(\frac{1}{N}\sum_{i=1}^{N}((x_i - \hat{x_i})^T u_j)^2) \\
&= \sum_{j=q+1}^{p} u_j^T S u_j
\end{aligned}$$

所以

$$\begin{cases}
\hat{u_j} = \arg\min\limits_{u_j} \sum\limits_{j=q+1}^{p} u_j^T S u_j \\
\\
u_j^T u_j = 1
\end{cases}$$

由于$u_{q+1}, u_{q+2}, ..., u_p$线性无关，所以这个最优化问题我们可以分别求解,和上一节的内容一致，所以

$$\begin{aligned}J &= \sum_{j=q+1}^{p} u_j^T S u_j \\ &= \sum_{j=q+1}^{p} u_j^T \lambda_j u_j \\ &= \sum_{j=q+1}^{p} \lambda_j\end{aligned}$$

我们选取那些小的特征值及其对应的特征向量，使得J最小。

SVD Perspective

前两节实际上都是在对样本的方差矩阵进行特征值分解：

$$S = GKG^T, G^TG = I, K = \Lambda(\lambda_1, \lambda_2, ..., \lambda_p)$$

这一节我们从样本数据着手，首先我们对数据进行中心化：

$$x_i := x_i - \overline{x}$$

矩阵表示即：

$$X := HX$$

然后，我们对HX进行奇异值分解：

$$HX = U\Sigma V^T$$

其中：$U^TU = I, V^TV = VV^T = I, \Sigma \in \Lambda$
那么

$$S_{p*p} = X^THX = X^TH^THX = V\Sigma U^TU\Sigma V^T = V\Sigma^2V^T$$

又因为$V^TV = I$，所以

$$
\begin{cases}
V = G \\
\\
\Sigma^2 = K
\end{cases}
$$

同样的：

$$T_{N*N} = HXX^TX = U\Sigma^2U^T$$

由此我们发现，S,T有相同的特征值，我们对S进行特征分解可以得到方向（主成分），然后$HX \cdot V$可以得到坐标；如果我们对T进行特征分解，我们可以直接得到坐标。（主坐标分析PCoA）

$$
\begin{aligned}
HXV = U\Sigma V^TV = U\Sigma \\
TU\Sigma = U\Sigma^2U^TU\Sigma = U\Sigma \cdot \Sigma^2
\end{aligned}
$$

Σ^2是T的特征值矩阵，所以$U\Sigma$是T的特征向量矩阵，也是坐标矩阵。