PCA原理推导

PCA使用方差来衡量新变量的信息量大小，按照方差大小排序依次为第一主成分、第二主成分等。下面对PCA原理进行简单推导。

假设原始数据为m维随机变量 $X=(X_1,X_2,\cdots X_m)^T$，其均值向量 $\mu=E(X)=(\mu_1,\mu_2,\cdots\mu_m)^T$，协方差矩阵为：

$$\Sigma=\mathrm{cov}(X,X)=E[(X-\mu)(X-\mu)^T]$$

由m维随机变量X到m维随机变量 $y=(y_1,y_2,\cdots y_m)^T$ 的线性变换：

$$y_i=\alpha_i^T X=\alpha_{1i}X_1+\alpha_{2i}X_2+\cdots+\alpha_{mi}X_m$$

其中 $\alpha_i^T=(\alpha_{1i},\alpha_{2i},\cdots\alpha_{mi})$。

经过线性变换后的随机变量 y_i 的均值、方差和协方差统计量表示为：

$$E(y_i)=\alpha_i^T\mu_i,\quad i=1,2,\cdots,m$$

$$\mathrm{var}(y_i)=\alpha_i^T\Sigma\alpha_i,\quad i=1,2,\cdots,m$$

$$\mathrm{cov}(y_i,y_j)=\alpha_i^T\Sigma\alpha_j,\quad i,j=1,2,\cdots,m$$

当随机变量X到随机变量y的线性变换满足如下条件时，变换后的 $y_1,y_2,\cdots y_m$ 分别为随机变量X的第一主成分、第二主成分，…，第m主成分。

(1) 线性变换的系数向量 α_i^T 为单位向量，有 $\alpha_i^T\alpha_i=1$，$i=1,2,\cdots,m$。

(2) 线性变换后的变量 y_i 与 y_j 线性无关，即 $\mathrm{cov}(y_i,y_j)\neq 0\ (i\neq j)$。

(3) 变量 y_1 是随机变量X所有线性变换中方差最大的，y_2 是 y_1 无关的所有线性变换中方差最大的。 方差大 → 坐标轴上打得散

上述条件给出了求解主成分的基本方法。根据优化目标和约束条件，我们可以用拉格朗日乘子法来求解主成分。下面以第一主成分为例求解推导：

$$\max\ \alpha_1^T\Sigma\alpha_1$$
$$\text{s.t.}\ \alpha_1^T\alpha_1=1$$

定义的拉格朗日函数如下：

$$L=\alpha_1^T\Sigma\alpha_1-\lambda(\alpha_1^T\alpha_1-1)$$

对α_1求导并令其为0，有：

$$\frac{\partial L}{\partial \alpha_1} = \Sigma\alpha_1 - \lambda\alpha_1 = 0$$

λ为Σ的特征值，α_1为对应的单位特征向量。假设α_1是Σ最大特征值λ_1对应的单位特征向量，那么α_1和λ_1均为上述优化问题的最优解。

$$var(\alpha_1^T x) = \alpha_1^T \Sigma \alpha_1 = \lambda_1$$

同样的方法来求解第k主成分，第k主成分的方差的第k个特征值为：

$$var(\alpha_k^T x) = \alpha_k^T \Sigma \alpha_k = \lambda_k, \quad k=1,2,\cdots m$$

最后梳理一下PCA的计算过程：

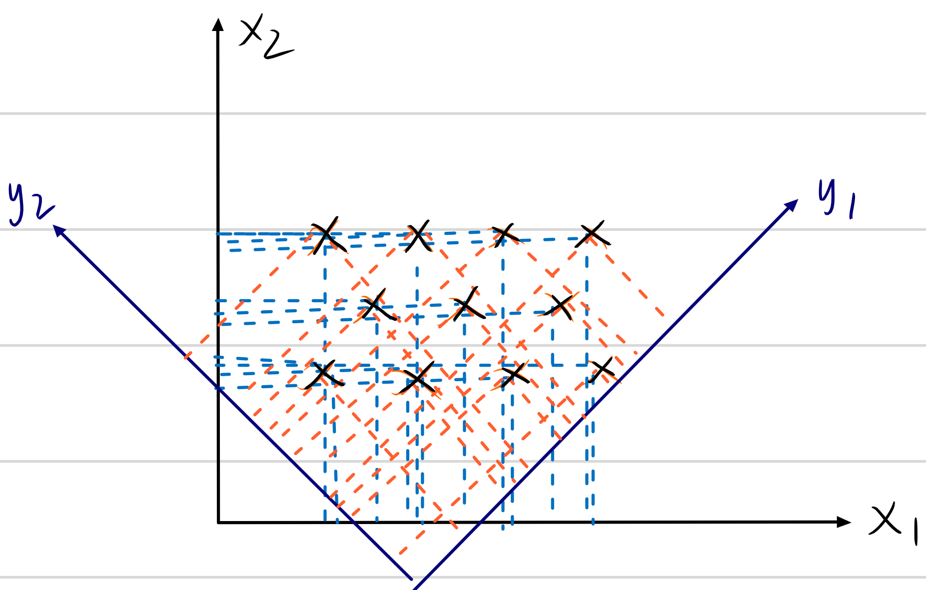
(1) 对m行n列的数据X按照均值为0，方差为1进行标准化处理。

(2) 计算标准化后的X的协方差矩阵 $C=\frac{1}{m}XX^T$。

(3) 计算协方差矩阵C的特征值和对应的特征向量。

(4) 将特征向量按照对应特征值大小排列成矩阵，取前k行构成矩阵P.

(5) 计算 $Y=PX$ 即可得到经过PCA降维后的k维数据。



PCA可以解决维度灾难问题，是对原始特征空间的重构。