

PCA 和 KPCA:

问 1: PCA 和 KPCA 具体操作流程。

x_i 表示第 i 个样本，为一个 $n \times 1$ 维列向量。 X 的每一行为一个样本的转置，为 $m \times n$ 的矩阵（共有 m 个样本）。

PCA:

1、特征中心化，即每一维数据都减去该维均值，除以该维标准差

$$\mu = \frac{1}{m} \sum_{i=1}^m x_i, \quad \sigma = \left(\frac{1}{m-1} \sum_{i=1}^m (x_i - \mu)^2 \right)^{\frac{1}{2}}$$
$$x_i = \frac{x_i - \mu}{\sigma}$$

2、计算协方差矩阵（可向量化处理）

$$C = \frac{1}{m} \sum_{i=1}^m x_i x_i^T, \quad C = \frac{1}{m} X^T X$$

3、计算协方差矩阵的特征值和特征向量，SVD 分解（svd() 已经排序）

$$C = U \Lambda U^T$$

4、从大到小选取若干特征值对应的特征向量，并映射得到新样本（可向量化处理）

$$U_k = [u_1, u_2, \dots, u_k], \quad z_i = U_k^T x_i, \quad Z = XU_k$$

5、可以使用映射样本复原原始样本

$$x'_i = U_k z_i, \quad X' = ZU_k^T$$

KPCA:

1、利用核函数 $K(x_i, x_j)$ 计算核矩阵 K ，同样需要保证映射后的向量均值为零，因此

$$K_{ij} = K(x_i, x_j)$$

$$\tilde{K} = K - I_m K - K I_m + I_m K I_m$$

其中 I_m 为每个元素都为 $1/m$ 的矩阵。

3、对核矩阵 \tilde{K} 进行 SVD 分解，选取前 k 个特征值和特征向量

$$\tilde{K} = U \Lambda U^T$$

$$m\lambda_i\alpha_i = \tilde{K}\alpha_i$$

4、对每个特征向量除以对应特征值的平方，对特征向量进行单位化

$$\alpha_i = \frac{\alpha_i}{\sqrt{\lambda_i}}$$

5、求得映射矩阵进行降维

$$U_k = [\alpha_1, \alpha_2, \dots, \alpha_k], \quad z_i = U_k^T \tilde{K}(i,:), \quad Z = \tilde{K}U_k$$

6、获得新数据 Y ，计算 K^Y ，并映射

$$K_{ij}^Y = K(y_i, x_j), \quad z_i = U_k^T K^Y(i,:), \quad Z^Y = K^Y U_k$$

7、关于 KPCA 数据的恢复，比较复杂请参见参考文档。

问 2：PCA 在哪一步引入了内积？KPCA 的核函数体现在哪里？

假定投影函数为 $\phi(\bullet)$ ，与 PCA 一样首先必须保证投影之后的样本均值为零

$$\frac{1}{m} \sum_{i=1}^m \phi(x_i) = 0$$

对应映射之后样本的协方差矩阵

$$C = \frac{1}{m} \sum_{i=1}^m \phi(x_i) \phi(x_i)^T$$

相应其特征值和特征向量应满足

$$Cv_k = \lambda_k v_k$$

将上面两式结合

$$\frac{1}{m} \sum_{i=1}^m \phi(x_i) \phi(x_i)^T v_k = \lambda_k v_k$$

而由于映射空间是由 $\phi(x_i)$ 所张成的空间，因此 v_k 能够用 $\phi(x_i)$ 线性表示。

$$v_k = \sum_{j=1}^m \alpha_k^j \phi(x_j)$$

将上式代入

$$\frac{1}{m} \sum_{i=1}^m \phi(x_i) \phi(x_i)^T \sum_{j=1}^m \alpha_k^j \phi(x_j) = \lambda_k \sum_{i=1}^m \alpha_k^i \phi(x_i)$$

此时两边同时乘以 $\phi(x_i)^T$ ，并直接核函数 $K_{ij} = \phi(x_i)^T \phi(x_j)$

$$\frac{1}{m} \sum_{i=1}^m K_{li} \sum_{j=1}^m \alpha_k^j K_{ij} = \lambda_k \sum_{i=1}^m \alpha_k^i K_{li}$$

其等价于

$$m \lambda_k K \alpha_k = K^2 \alpha_k \Rightarrow m \lambda_k \alpha_k = K \alpha_k$$

此时 $\phi(x_i)$ 投影的第 k 个元素

$$z_k(x_i) = \phi(x_i)^T v_k = \sum_{j=1}^m \alpha_k^j K_{ij}$$

由于映射之后的样本均值不一定为零，因此需要归一化

$$\phi(x_i) = \phi(x_i) - \frac{1}{m} \sum_{i=1}^m \phi(x_i)$$

对应核矩阵

$$\tilde{K} = K - I_m K - K I_m + I_m K I_m$$

由于 v_k 才是投影空间的特征向量，而 α_k 只是表示参数，因此需要对 α_k 进行归一化

$$\begin{aligned} 1 &= v_n^T v_n = \sum_{i=1}^m \alpha_n^i \phi(x_i)^T \sum_{j=1}^m \alpha_n^j \phi(x_j) \\ &= \sum_{i,j=1}^m \alpha_n^i \alpha_n^j \phi(x_i)^T \phi(x_j) \\ &= \sum_{i,j=1}^m \alpha_n^i \alpha_n^j K_{ij} \\ &= \alpha_n^T K \alpha_n \\ &= \lambda_n \alpha_n^T \alpha_n \end{aligned}$$

因此需要对 α_i 进行如下归一化

$$\alpha_i = \frac{\alpha_i}{\sqrt{\lambda_i}}$$

- 1、Principal Component Analysis (PCA) 是最常用的**线性**降维方法，它的目标是通过某种线性投影，将高维的数据映射到低维的空间中表示，并期望在所投影的维度上数据的**方差最大**，以此使用较少的数据维度，同时保留住较多的原数据点的特性。
- 2、LDA: Linear Discriminant Analysis (也有叫做 Fisher Linear Discriminant)是一种有监督的线性降维算法。与 PCA 保持数据信息不同，LDA 是为了使得降维后的数据点尽可能地容易被区分。
- 3、