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 $m\times n$ 为的矩阵（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

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

$$x_i=\frac{x_i-\mu}{\sigma}$$

2、计算协方差矩阵（可向量化处理）

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

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、从大到小选取诺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）

$$U_k=\left[u_1,u_2,\cdots,u_k\right]\text{，}\quad z_i=U_k^T x_i\text{，}\quad Z=XU_k$$

5、可以使用映射样本复原原始样本

$$x_i'=U_k z_i\text{，}\quad X'=ZU_k^T$$

KPCA:

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

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

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

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

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

6、获得新数据Y，计算K^Y，并映射

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

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

$$C v_k = \lambda_k v_k$$

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

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

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

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

其等价于

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

此时$\phi\left(x_i\right)$投影的第k个元素

$$z_k(x_i)=\phi\left(x_i\right)^Tv_k=\sum_{j=1}^{m}\alpha_k^iK_{ij}$$

由于映射之后的样本均值不一定为零，因此需要归一化

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

对应核矩阵

$$\tilde{K}=K-I_mK-KI_m+I_mKI_m$$

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

$$\begin{aligned}1=v_i^Tv_j&=\sum_{s=1}^{m}\alpha_i^s\phi\left(x_s\right)^T\sum_{t=1}^{m}\alpha_j^t\phi\left(x_t\right)\\&=\sum_{s,t=1}^{m}\alpha_i^s\alpha_j^t\phi\left(x_s\right)^T\phi\left(x_t\right)\\&=\sum_{s,t=1}^{m}\alpha_i^s\alpha_j^tK_{st}\\&=\alpha_i^TK\alpha_j\\&=\lambda_k\alpha_i^T\alpha_j\end{aligned}$$

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、