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）

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

5、可以使用映射样本复原原始样本

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

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

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

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

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

$$U_k = \left[\alpha^1, \alpha^2, \cdots, \alpha^k\right], \quad z_i = U_k^T \tilde{K}\left(i,:\right), \quad Z = \tilde{K}U_k$$

6、获得新数据Y，计算K^Y，并映射

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

7、关于 KPCA 数据的恢复，由于只能获得映射后映射空间中的值，因此通常比较复杂。详情请参见参考文档。

问 2：PCA 在哪一步引入了内积？KPCA 的核函数体现在哪里？

PCA 的协方差矩阵，其中$\frac{1}{m}\sum_{i=1}^{m} x_i = 0$

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

对应特征值和特征向量满足

$$Cv = \lambda v$$

从而得到

$$\lambda v = Cv = \frac{1}{m}\sum_{i=1}^{m}\left\langle x_i, v\right\rangle x_i$$

两边同时左乘x_i^T

$$\lambda\left\langle x_i, v\right\rangle = \left\langle x_i, Cv\right\rangle$$

而$\left\langle x_i, v\right\rangle$就是需要的在$v$上的投影。

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

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

对应映射之后样本的协方差矩阵

$$C=\frac{1}{m}\sum_{j=1}^{m}\phi\left(x_j\right)\phi\left(x_j\right)^T$$

相应其特征值和特征向量应满足

$$Cv=\lambda v$$

同理可得

$$\lambda\left\langle\phi\left(x_n\right),v\right\rangle=\left\langle\phi\left(x_n\right),Cv\right\rangle$$

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

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

将上式代入（注意求和结合是向量还是矩阵能够更好的理解求和次序的交换）

$$\lambda\sum_{i=1}^{m}\alpha_i\left\langle\phi\left(x_n\right),\phi\left(x_i\right)\right\rangle=\frac{1}{m}\sum_{i=1}^{m}\alpha_i\left\langle\phi\left(x_n\right),\sum_{j=1}^{m}\phi\left(x_j\right)\left\langle\phi\left(x_j\right)\phi\left(x_i\right)\right\rangle\right\rangle$$

并假定核函数$K_{ij}=\left\langle\phi\left(x_i\right),\phi\left(x_j\right)\right\rangle$

$$\lambda\sum_{i=1}^{m}\alpha_iK_{ni}=\frac{1}{m}\sum_{i=1}^{m}\alpha_i\sum_{j=1}^{m}K_{nj}K_{ji}$$

其等价于，由于K半正定

$$m\lambda K\alpha=K^2\alpha\Rightarrow m\lambda\alpha=K\alpha$$

如果λ_k对应参数为α^k。此时$\phi\left(x\right)$映射后的第k个元素

$$z_k(x)=\left\langle\phi\left(x\right),v_k\right\rangle=\sum_{i=1}^{m}\alpha_i^k\left\langle\phi\left(x_i\right),\phi\left(x\right)\right\rangle$$

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，因此还需要对α^k进行归一化

$$
\begin{aligned}
1=\langle v_n,v_n\rangle &= \sum_{i=1}^{m}\alpha_i^n\alpha_j^n\left\langle \phi(x_i),\phi(x_j)\right\rangle \\
&= \sum_{i,j=1}^{m}\alpha_i^n\alpha_j^n K_{ij} \\
&= \sum_{i=1}^{m}\alpha_i^n\sum_{j=1}^{m}\alpha_j^n K_{ij} \\
&= \left\langle \alpha^n, K\alpha^n\right\rangle \\
&= \lambda_n\left\langle \alpha^n,\alpha^n\right\rangle
\end{aligned}
$$

因此需要对α^i进行如下归一化

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

问 3：理解 KPCA。

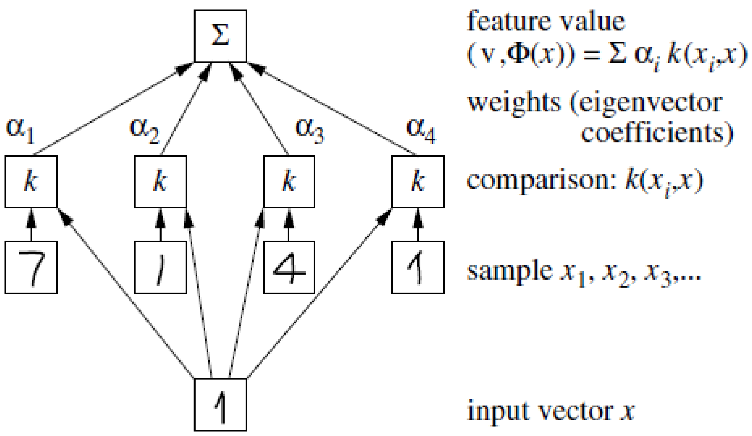

Figure 14.2 Feature extractor constructed using Kernel PCA (cf. (14.16)). In the first layer, the input vector is compared to the sample via a kernel function, chosen a priori (e.g. polynomial, Gaussian, or sigmoid). The outputs are then linearly combined using weights, which are found by solving an eigenvector problem. As shown in the text, the function of the network can be thought of as the projection onto an eigenvector of a covariance matrix in a high-dimensional feature space. As a function on input space, it is nonlinear.

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、