第 8 章补充材料

● 矢量与坐标系

特征矢量一般都是以坐标的形式表示的，矢量的每一个分量是它在相应坐标轴上的投影。坐标系可以用坐标原点 O 和一组代表各个坐标轴的"基矢量"$\{\mathbf{e}_1,\mathbf{e}_2,\cdots,\mathbf{e}_d\}$来表示，如果基矢量之间满足如下关系：

$$\mathbf{e}_i^t\mathbf{e}_j=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases} \tag{1}$$

则称这组基矢量构成了一个标准正交系，所谓"标准"是指每个矢量的长度均为 1，所谓"正交"是指任意两个矢量之间是正交的。这样构成的坐标系就是我们所熟悉的直角坐标系。

如图 1 所示，在一个直角坐标系下，矢量$\mathbf{x}=(x_1,\cdots,x_d)^t$可以表示为基矢量的线性组合：

$$\mathbf{x}=x_1\mathbf{e}_1+x_2\mathbf{e}_2+\cdots+x_d\mathbf{e}_d=\sum_{i=1}^{d}x_i\mathbf{e}_i \tag{2}$$

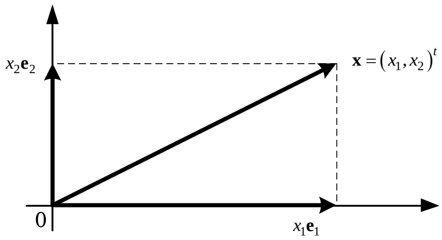

图 1 矢量和直角坐标系

很显然，根据（1）和（2）式，我们可以通过计算矢量$\mathbf{x}$与基矢量$\mathbf{e}_j$的内积来得到相应的坐标分量：

$$\mathbf{x}^t\mathbf{e}_j=\sum_{i=1}^{d}x_i\mathbf{e}_i^t\mathbf{e}_j=x_j \tag{3}$$

● 主成分分析算法的推导

我们知道，样本集合中的每一个样本都对应着特征空间中的一个点，同样的一个样本点在不同坐标系下对应着不同的矢量，例如图 2 中某个样本对应着特征空间中的点 A，在以 O 为坐标原点$\{\mathbf{v}_1,\mathbf{v}_2\}$为基矢量的原坐标系下 A 点对应的矢量是$\mathbf{x}$，而在 O’为原点$\{\mathbf{e}_1,\mathbf{e}_2\}$为基矢量的新坐标系下对应的矢量是$\mathbf{x}'$，如果新坐标系的原点 O'在原坐标系下对应的矢量是$\boldsymbol{\mu}$，显然有如下关系：

$$\mathbf{x}=\boldsymbol{\mu}+\mathbf{x}' \tag{1}$$

这个关系虽然是从 2 维空间中得到的，在高维空间中同样成立。

分别将矢量$\mathbf{x}$和$\mathbf{x}'$写成两个坐标系下的坐标形式：$\mathbf{x}=\left(x_1,\cdots,x_d\right)^t$，$\mathbf{x}'=\left(a_1,\cdots,a_d\right)^t$，（1）式可以表示为：

$$\mathbf{x}=\boldsymbol{\mu}+\sum_{i=1}^{d}a_i\mathbf{e}_i \tag{2}$$

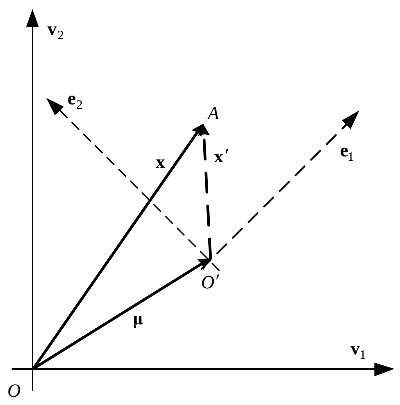

图 2 同一个样本在不同坐标系下的表示

在新坐标系下，矢量$\mathbf{x}'$的元素可以由原坐标系下的矢量$\mathbf{x}$以及$\boldsymbol{\mu}$和基矢量$\{\mathbf{e}_1,\cdots,\mathbf{e}_d\}$计算得到：

$$a_i=\mathbf{e}_i^t\left(\mathbf{x}-\boldsymbol{\mu}\right),\quad i=1,\cdots,d \tag{3}$$

同样也可以根据（2）式由新坐标下的矢量$\mathbf{x}'$来恢复原矢量$\mathbf{x}$，不会存在任何误差。然而如果我们只保留新坐标系下$d'<d$个元素，然后用保留的d'个元素来恢复原坐标系下的d维矢量：

$$\hat{\mathbf{x}}=\boldsymbol{\mu}+\sum_{i=1}^{d'}a_i\mathbf{e}_i \tag{4}$$

显然$\hat{\mathbf{x}}$只是对$\mathbf{x}$的近似，用$\hat{\mathbf{x}}$来代替$\mathbf{x}$就会出现一定的误差，误差的大小同新坐标系的位置、基矢量的方向以及保留哪些特征有关。

在主成分分析方法中，新坐标系的原点选择在训练样本集$D=\{\mathbf{x}_1,\cdots,\mathbf{x}_n\}$的均值矢量$\boldsymbol{\mu}$上，然后寻找一组最优的基矢量$\{\mathbf{e}_1,\cdots,\mathbf{e}_d\}$，使得在只保留前$d'$个元素的条件下，由新的坐标根据（4）式恢复样本集$D$的均方误差最小，即求解如下的优化问题：

$$\min_{\mathbf{e}_1,\cdots,\mathbf{e}_d} J\left(\mathbf{e}_1,\cdots,\mathbf{e}_d\right)=\frac{1}{n}\sum_{k=1}^{n}\left\|\mathbf{x}_k-\hat{\mathbf{x}}_k\right\|^2 \tag{5}$$

其中$\hat{\mathbf{x}}_k$是根据（3）式将$\mathbf{x}_k$由原坐标系变换到新坐标系下，然后再根据（4）式只使用前d'个特征恢复的近似矢量。如果用a_{ki}表示第k个样本在新坐标系下的第i维特征，由（2）式和（4）式可以得到：

$$\mathbf{x}_k-\hat{\mathbf{x}}_k=\sum_{i=d'+1}^{d}a_{ki}\mathbf{e}_i \tag{6}$$

代入到（5）式：

$$\begin{aligned}
J\left(\mathbf{e}_1,\cdots,\mathbf{e}_d\right)&=\frac{1}{n}\sum_{k=1}^{n}\left\|\mathbf{x}_k-\hat{\mathbf{x}}_k\right\|^2=\frac{1}{n}\sum_{k=1}^{n}\left\|\sum_{i=d'+1}^{d}a_{ki}\mathbf{e}_i\right\|^2\\
&=\frac{1}{n}\sum_{k=1}^{n}\left(\sum_{i=d'+1}^{d}a_{ki}\mathbf{e}_i\right)^t\left(\sum_{i=d'+1}^{d}a_{ki}\mathbf{e}_i\right)\\
&=\frac{1}{n}\sum_{k=1}^{n}\sum_{i=d'+1}^{d}a_{ki}^2\\
&=\frac{1}{n}\sum_{k=1}^{n}\sum_{i=d'+1}^{d}\left[\mathbf{e}_i^t\left(\mathbf{x}_k-\boldsymbol{\mu}\right)\right]\left[\mathbf{e}_i^t\left(\mathbf{x}_k-\boldsymbol{\mu}\right)\right]^t\\
&=\sum_{i=d'+1}^{d}\mathbf{e}_i^t\left[\frac{1}{n}\sum_{k=1}^{n}\left(\mathbf{x}_k-\boldsymbol{\mu}\right)\left(\mathbf{x}_k-\boldsymbol{\mu}\right)^t\right]\mathbf{e}_i
\end{aligned}$$

其中第2行到第3行利用了$\left\{\mathbf{e}_1,\cdots,\mathbf{e}_d\right\}$是新坐标系的基矢量，因此构成了一个标准正交系：

$$\mathbf{e}_i^t\mathbf{e}_j=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}，\quad i,j=1,\cdots,d$$

而第3行到第4行则是基于如下事实：a_{ki}是一个标量，它的转置与其自身相等，并且有（3）式成立，因此$a_{ki}=\mathbf{e}_i^t\left(\mathbf{x}_k-\boldsymbol{\mu}\right)=\left[\mathbf{e}_i^t\left(\mathbf{x}_k-\boldsymbol{\mu}\right)\right]^t$。如果定义矩阵：

$$\boldsymbol{\Sigma}=\frac{1}{n}\sum_{k=1}^{n}\left(\mathbf{x}_k-\boldsymbol{\mu}\right)\left(\mathbf{x}_k-\boldsymbol{\mu}\right)^t$$

恰好是样本集D的协方差矩阵，则（5）式的优化问题变为：

$$\min_{\mathbf{e}_1,\cdots,\mathbf{e}_d} J\left(\mathbf{e}_1,\cdots,\mathbf{e}_d\right)=\sum_{i=d'+1}^{d}\mathbf{e}_i^t\boldsymbol{\Sigma}\mathbf{e}_i \tag{7}$$

仔细观察（7）式会发现，直接求解这个优化问题是没有意义的。由于$\boldsymbol{\Sigma}$是半正定矩阵，因此当$\mathbf{e}_1,\cdots,\mathbf{e}_d=\mathbf{0}$时取得最小值0，显然零矢量并不能作为基矢量，导致这样结果的原因在于优化（7）式时没有约束$\mathbf{e}_1,\cdots,\mathbf{e}_d$的长度。主成分分析在求解新坐标系的基矢量时优化的是如下的约束问题：

$$\min_{\mathbf{e}_1,\cdots,\mathbf{e}_d} J\left(\mathbf{e}_1,\cdots,\mathbf{e}_d\right)=\sum_{i=d'+1}^{d}\mathbf{e}_i^t\boldsymbol{\Sigma}\mathbf{e}_i \tag{8}$$

约束：

$$\left\|\mathbf{e}_i\right\|^2=1，\quad i=1,\cdots,d$$

有约束优化问题可以通过构造 Lagrange 函数转化为无约束问题(参见最优化方法 VI):

$$L(\mathbf{e}_1,\cdots,\mathbf{e}_d,\lambda_1,\cdots,\lambda_d)=\sum_{i=d'+1}^{d}\mathbf{e}_i^t\boldsymbol{\Sigma}\mathbf{e}_i-\sum_{i=d'+1}^{d}\lambda_i\left(\mathbf{e}_i^t\mathbf{e}_i-1\right) \tag{9}$$

对每一个基矢量 $\mathbf{e}_j$ 求导数:

$$\frac{\partial L(\mathbf{e}_1,\cdots,\mathbf{e}_d,\lambda_1,\cdots,\lambda_d)}{\partial\mathbf{e}_j}=2\boldsymbol{\Sigma}\mathbf{e}_j-2\lambda_j\mathbf{e}_j=0 \tag{10}$$

其中利用到了 $\boldsymbol{\Sigma}$ 为对称矩阵的事实。由此得到优化问题（8）的解应满足:

$$\boldsymbol{\Sigma}\mathbf{e}_j=\lambda_j\mathbf{e}_j \tag{11}$$

显然，使得上式成立的 λ_j 和 $\mathbf{e}_j$ 分别为矩阵 $\boldsymbol{\Sigma}$ 的特征值和对应的特征矢量。由此我们可以得到这样的结论：如果我们希望将一个样本集合 D 中的 d 维特征矢量在一个新的坐标系下只用 d' 个特征进行表示，那么应该将新坐标系的坐标原点放在 D 的均值 $\boldsymbol{\mu}$ 的位置，而以集合 D 的协方差矩阵的特征矢量 $\mathbf{e}_1,\cdots,\mathbf{e}_d$ 作为基矢量，这样可以保证只用保留的 d' 维特征恢复原矢量时均方误差最小。

通过这样的方式可以得到一个最优的新坐标系，注意到 $\boldsymbol{\Sigma}$ 是一个 $d\times d$ 的矩阵，存在 d 个特征值和特征矢量，现在的问题是我们只希望保留新坐标系中的 d' 个坐标，应该保留哪些坐标才能够保证恢复出的 d 维特征矢量的均方误差最小？回到优化问题（8），将（11）式代入优化函数:

$$J(\mathbf{e}_1,\cdots,\mathbf{e}_d)=\sum_{i=d'+1}^{d}\mathbf{e}_i^t\boldsymbol{\Sigma}\mathbf{e}_i=\sum_{i=d'+1}^{d}\lambda_i\mathbf{e}_i^t\mathbf{e}_i=\sum_{i=d'+1}^{d}\lambda_i \tag{12}$$

要使得 $J(\mathbf{e}_1,\cdots,\mathbf{e}_d)$ 最小，只需要选择 $\lambda_{d'+1},\cdots,\lambda_d$ 是 $\boldsymbol{\Sigma}$ 最小的 $d-d'$ 个特征值。这里需要注意一点，在整个推导过程中我们约定的是要保留新坐标系下前 d' 个特征，而放弃掉后面的 $d-d'$ 个特征，因此在新的坐标系下应该选择保留的是 $\boldsymbol{\Sigma}$ 最大的 d' 个特征值对应的特征矢量作为新坐标系的基矢量。

● 基于 Fisher 准则的可分性分析算法的推导

样本下面先从一种简单的情况入手来研究这个问题，我们将两个类别的样本向一条通过坐标原点的直线上投影，也就是用 1 维特征来表示 d 维矢量，希望在 1 维空间中两类样本的可分性最大。从图 3 可以看出在不同方向的直线上，两类样本的可分性是不同的，如果想要找到一个最优的投影直线方向，首先需要对 1 维空间中样本的可分性进行度量。

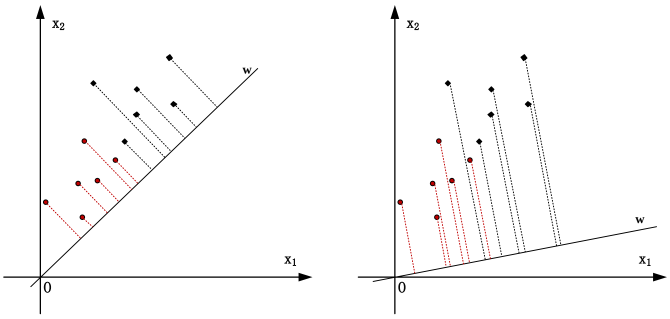

图 3 二维模式在一维空间的投影

假设两类问题的样本集为：$D_1=\left\{\mathbf{x}_1^{(1)},\cdots,\mathbf{x}_{n_1}^{(1)}\right\}$，$D_2=\left\{\mathbf{x}_1^{(2)},\cdots,\mathbf{x}_{n_2}^{(2)}\right\}$，投影直线的单位矢量为 $\mathbf{w}$，d 维空间的矢量 $\mathbf{x}$ 在这条直线上的投影为一个标量：

$$y=\mathbf{w}^t\mathbf{x} \tag{1}$$

两类样本集经过投影之后成为标量集：$D_1\to\mathcal{Y}_1=\left\{y_1^{(1)},\cdots,y_{n_1}^{(1)}\right\}$，$D_2\to\mathcal{Y}_2=\left\{y_1^{(2)},\cdots,y_{n_2}^{(2)}\right\}$。不同类样本的分散程度越大，同类样本的聚集程度越高则类别之间的可分性越强。在 1 维空间中可以用两个类别样本均值之差的平方 $(\tilde{\mu}_1-\tilde{\mu}_2)^2$ 来度量两类样本的分散程度；而样本类内的离散程度可以用样本的方差之和来度量 $\tilde{s}_1^2+\tilde{s}_2^2$。综合考虑类内的聚集程度和类间的分散程度，可以建立如下的 Fisher 准则：

$$J(\mathbf{w})=\frac{(\tilde{\mu}_1-\tilde{\mu}_2)^2}{\tilde{s}_1^2+\tilde{s}_2^2} \tag{2}$$

Fisher 准则函数的值越大，类别的可分性则越强。下面写出准则函数 $J(\mathbf{w})$ 关于投影直线方向矢量 $\mathbf{w}$ 的显式表达式，首先计算投影之后类别的均值 $\tilde{\mu}$：

$$\tilde{\mu}_i=\frac{1}{n_i}\sum_{y\in\mathcal{Y}_i}y=\frac{1}{n_i}\sum_{\mathbf{x}\in D_i}\mathbf{w}^t\mathbf{x}=\mathbf{w}^t\boldsymbol{\mu}_i,\quad i=1,2 \tag{3}$$

投影之后两类均值之差的平方可以表示为：

$$(\tilde{\mu}_1-\tilde{\mu}_2)^2=\left(\mathbf{w}^t\boldsymbol{\mu}_1-\mathbf{w}^t\boldsymbol{\mu}_2\right)^2=\mathbf{w}^t(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\mathbf{w}=\mathbf{w}^t\mathbf{S}_b\mathbf{w} \tag{4}$$

其中 $\mathbf{S}_b=(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t$ 是类间散布矩阵（假设两类的先验概率相等）。类似的：

$$
\begin{aligned}
\tilde{s}_i^2 &= \sum_{y\in\mathcal{Y}_i}\left(y-\tilde{\mu}_i\right)^2 \\
&= \sum_{\mathbf{x}\in D_i}\left(\mathbf{w}^t\mathbf{x}-\mathbf{w}^t\boldsymbol{\mu}_i\right)^2 \\
&= \sum_{\mathbf{x}\in D_i}\mathbf{w}^t\left(\mathbf{x}-\boldsymbol{\mu}_i\right)\left(\mathbf{x}-\boldsymbol{\mu}_i\right)^t\mathbf{w} \\
&= \mathbf{w}^t\mathbf{S}_i\mathbf{w}
\end{aligned}
$$

其中 $\mathbf{S}_i=\sum_{\mathbf{x}\in D_i}\left(\mathbf{x}-\boldsymbol{\mu}_i\right)\left(\mathbf{x}-\boldsymbol{\mu}_i\right)^t$ 类似于类内散布矩阵。总的方差：

$$\tilde{s}_1^2+\tilde{s}_2^2=\mathbf{w}^t\mathbf{S}_1\mathbf{w}+\mathbf{w}^t\mathbf{S}_2\mathbf{w}=\mathbf{w}^t\left(\mathbf{S}_1+\mathbf{S}_2\right)\mathbf{w}=\mathbf{w}^t\mathbf{S}_w\mathbf{w} \tag{5}$$

（4）、（5）式代入 Fisher 准则，可以得到如下优化问题：

$$\max_{\mathbf{w}} J\left(\mathbf{w}\right)=\frac{\left(\tilde{\mu}_1-\tilde{\mu}_2\right)^2}{\tilde{s}_1^2+\tilde{s}_2^2}=\frac{\mathbf{w}^t\mathbf{S}_b\mathbf{w}}{\mathbf{w}^t\mathbf{S}_w\mathbf{w}} \tag{6}$$

上式也被称为是 Rayleigh 商的优化问题。实际上这个问题存在着无穷多个解，因为如果 $\mathbf{w}^*$ 是一个最优解的话，对于任意的 $a\neq 0$，$a\mathbf{w}^*$ 同样是最优解。我们真正关心的是投影矢量 $\mathbf{w}$ 的方向，而不关心它的长度（可以规格化为单位矢量），因此可以通过适当调整 $\|\mathbf{w}\|$ 使得 Fisher 准则的分母 $\mathbf{w}^t\mathbf{S}_w\mathbf{w}$ 等于一个常数 C。这样我们就得到了一个有约束的优化问题：

$$\max_{\mathbf{w}} J\left(\mathbf{w}\right)=\mathbf{w}^t\mathbf{S}_b\mathbf{w} \tag{7}$$

约束：

$$\mathbf{w}^t\mathbf{S}_w\mathbf{w}=C$$

构造 Lagrange 函数转化为无约束优化：

$$L\left(\mathbf{w}\right)=\mathbf{w}^t\mathbf{S}_b\mathbf{w}-\lambda\left(\mathbf{w}^t\mathbf{S}_w\mathbf{w}-C\right) \tag{8}$$

对 $\mathbf{w}$ 求导：

$$\frac{\partial L\left(\mathbf{w}\right)}{\partial\mathbf{w}}=2\mathbf{S}_b\mathbf{w}-2\lambda\mathbf{S}_w\mathbf{w}=0$$

因此有：

$$\mathbf{S}_b\mathbf{w}=\lambda\mathbf{S}_w\mathbf{w} \tag{9}$$

满足上式的 λ 和 $\mathbf{w}$ 称为关于 $\mathbf{S}_b$ 和 $\mathbf{S}_w$ 的广义特征值和特征矢量，如果 $\mathbf{S}_w$ 的逆矩阵存在的话，有：

$$\mathbf{S}_w^{-1}\mathbf{S}_b\mathbf{w}=\lambda\mathbf{w} \tag{10}$$

λ 和 $\mathbf{w}$ 分别是矩阵 $\mathbf{S}_w^{-1}\mathbf{S}_b$ 的特征值和对应的特征矢量。与主成分分析一样，$\mathbf{S}_w^{-1}\mathbf{S}_b$ 是一个 $d\times d$ 的方阵，存在 d 个特征值和 d 个特征矢量，哪一个特征矢量使得 Fisher 准则取得最大值？将满足（9）式的一个特征矢量 $\mathbf{w}_i$ 代入（6）式：

$$J\left(\mathbf{w}\right)=\frac{\mathbf{w}_i^t\mathbf{S}_b\mathbf{w}_i}{\mathbf{w}_i^t\mathbf{S}_w\mathbf{w}_i}=\frac{\lambda_i\mathbf{w}^t\mathbf{S}_w\mathbf{w}_i}{\mathbf{w}_i^t\mathbf{S}_w\mathbf{w}_i}=\lambda_i \tag{11}$$

由此可见最大特征值对应的特征矢量是使得 Fisher 准则取得最大值的方向矢量。通过这样一个推导过程我们可以得出如下结论：两个类别的样本向一条直线上投影，当直线的方向矢量 $\mathbf{w}$ 为矩阵 $\mathbf{S}_w^{-1}\mathbf{S}_b$ 最大特征值对应的特征矢量时，可以使得投影之后样本在 1 维空间中

具有最大的可分性。

下面将这个问题推广到 c 个类别的样本向 d' 个方向上的投影，即将 d 维特征降维为 d' 个特征，使得降维之后的样本具有最大的可分性。首先定义矩阵：

$$\mathbf{S}_w = \sum_{i=1}^{c} \mathbf{S}_i \text{ ，} \quad \mathbf{S}_b = \sum_{i=1}^{c} n_i \left(\boldsymbol{\mu}_i - \boldsymbol{\mu}\right)\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}\right)^t \tag{12}$$

其中 $\boldsymbol{\mu}$ 为所有样本的均值矢量。可以证明使得 Fisher 准则最大的 d' 个投影矢量是对应于矩阵 $\mathbf{S}_w^{-1}\mathbf{S}_b$ 最大 d' 个特征值的特征矢量。

● KPCA 方法的推导

首先证明该样本集合协方差矩阵的特征向量处于样本所张成的空间，即：$v \in span\{x_1, \cdots, x_m\}$，设有样本集合：$X = \{\mathbf{x}_1, \cdots, \mathbf{x}_m\}$，假设样本的均值为 0，即：$\sum_{i=1}^{m} \mathbf{x}_i = 0$，。

[证明]

样本集合的协方差矩阵为：$C = \frac{1}{m}\sum_{j=1}^{m} x_j \cdot x_j^t$，令 λ 为 C 的特征值，$\mathbf{v}$ 为对应特征矢量，则有：

$$\lambda \mathbf{v} = C\mathbf{v} = \left(\frac{1}{m}\sum_{j=1}^{m} \mathbf{x}_j \cdot \mathbf{x}_j^t\right)\mathbf{v} = \frac{1}{m}\sum_{j=1}^{m}\left(\mathbf{x}_j^t \mathbf{v}\right)\mathbf{x}_j \tag{1}$$

即：

$$\mathbf{v} = \frac{1}{m\lambda}\sum_{j=1}^{m}\left(\mathbf{x}_j^t \mathbf{v}\right)\mathbf{x}_j$$

因此：$v \in span\{x_1, \cdots, x_m\}$。　　　　证毕

下面来推导特征空间中协方差矩阵特征值和特征向量的求解方法。设有样本集合：$X = \{\mathbf{x}_1, \cdots, \mathbf{x}_m\}$，取非线性映射：$\phi : R^d \to \mathcal{F}$，$R^d$ 为样本所处的 d 维欧氏空间，称为输入空间，$\mathcal{F}$ 为一个 Hilbert 空间，称为特征空间，样本在特征空间中的内积可以用一个核函数来计算：$k(\mathbf{x}, \mathbf{y}) = \left(\phi(\mathbf{x})\right)^t \phi(\mathbf{y})$。假设样本集合在特征空间中的均值为 0，即：$\sum_{i=1}^{m}\phi(\mathbf{x}_i) = 0$。

样本在特征空间中的协方差矩阵为：

$$\overline{C} = \frac{1}{m}\sum_{i=1}^{m}\phi(\mathbf{x}_i)\phi^t(\mathbf{x}_i) \tag{2}$$

令：λ 为 $\overline{C}$ 的特征值，$\mathbf{V}$ 为对应的特征矢量，则有：

$$\lambda \mathbf{V} = \overline{C}\mathbf{V} \tag{3}$$

利用之前证明的结果，特征矢量 $\mathbf{V} \in span\{\phi(\mathbf{x}_1), \cdots, \phi(\mathbf{x}_m)\}$，即存在 $\alpha_1, \cdots, \alpha_m$，使得：

$$\mathbf{V} = \sum_{i=1}^{m}\alpha_i \phi(\mathbf{x}_i) \tag{4}$$

(4)式左乘 $\phi^t(\mathbf{x}_j)$，则有：

$$\lambda\left[\phi^t(\mathbf{x}_j)\cdot \mathbf{V}\right] = \phi^t(\mathbf{x}_j)\cdot \overline{C}\mathbf{V}, \quad j = 1, \cdots, m \tag{5}$$

将(4)式代入(5)式左端：

$$\begin{aligned}\lambda\left[\phi^t(\mathbf{x}_j)\cdot \mathbf{V}\right] &= \lambda\left(\sum_{i=1}^{m}\alpha_i \phi^t(\mathbf{x}_j)\phi(\mathbf{x}_i)\right) \\ &= \lambda\left(\sum_{i=1}^{m}\alpha_i k_{ji}\right) \\ &= \lambda\left(\mathbf{K}\boldsymbol{\alpha}\right)_j\end{aligned} \tag{6}$$

其中：

$$\mathbf{K}=\left[k\left(\mathbf{x}_i,\mathbf{x}_j\right)\right]_{m\times m}=\left(\phi^t\left(\mathbf{x}_i\right)\phi\left(\mathbf{x}_j\right)\right)_{m\times m}$$

$$\boldsymbol{\alpha}=\left(\alpha_1,\cdots,\alpha_m\right)^t$$

$\left(\mathbf{K}\boldsymbol{\alpha}\right)_j$ 为矢量 $\mathbf{K}\boldsymbol{\alpha}$ 的第 j 个元素.

将(2)和(4)式代入(5)式的右端：

$$\begin{aligned}\phi^t\left(\mathbf{x}_j\right)\cdot\overline{C}\mathbf{V}&=\phi^t\left(\mathbf{x}_j\right)\cdot\frac{1}{m}\sum_{l=1}^{m}\phi\left(\mathbf{x}_l\right)\phi^t\left(\mathbf{x}_l\right)\cdot\sum_{i=1}^{m}\alpha_i\,\phi\left(\mathbf{x}_i\right)\\&=\frac{1}{m}\sum_{l=1}^{m}\sum_{i=1}^{m}\alpha_i\left(\phi^t\left(\mathbf{x}_j\right)\phi\left(\mathbf{x}_l\right)\cdot\phi^t\left(\mathbf{x}_l\right)\phi\left(\mathbf{x}_i\right)\right)\\&=\frac{1}{m}\sum_{l=1}^{m}\sum_{i=1}^{m}\alpha_i k_{jl}k_{li}\\&=\frac{1}{m}\left(\mathbf{K}^2\boldsymbol{\alpha}\right)_j\end{aligned}\tag{7}$$

(6)式(7)式相等，则有：

$$\lambda\mathbf{K}\boldsymbol{\alpha}=\frac{1}{m}\mathbf{K}^2\boldsymbol{\alpha}\tag{8}$$

因此，求取 $\overline{C}$ 的特征值和特征向量的问题可以转化为如下特征值问题：

$$m\lambda\boldsymbol{\alpha}=\mathbf{K}\boldsymbol{\alpha}\tag{9}$$

用 λ 代替 $m\lambda$，即为：

$$\lambda\boldsymbol{\alpha}=\mathbf{K}\boldsymbol{\alpha}\tag{10}$$

令：$\lambda^1,\cdots,\lambda^m$ 为矩阵 $\mathbf{K}$ 的特征值，$\boldsymbol{\alpha}^1,\cdots,\boldsymbol{\alpha}^m$ 为特征矢量，则对应的 $\overline{C}$ 的第 i 个特征值和特征矢量为：

$$\lambda_i=\frac{\lambda^i}{m}\,,\quad \mathbf{V}_i=\sum_{j=1}^{m}\alpha_j^i\phi\left(\mathbf{x}_j\right)\tag{11}$$

其中 α_j^i 为 $\boldsymbol{\alpha}^i$ 的第 j 个元素。

规范化，将 $\mathbf{V}_i$ 转化为单位矢量。

$$\begin{aligned}\left(\mathbf{V}_i\right)^t\cdot\mathbf{V}_i&=\sum_{j,l=1}^{m}\alpha_j^i\alpha_l^i\left(\phi^t\left(\mathbf{x}_j\right)\cdot\phi\left(\mathbf{x}_l\right)\right)\\&=\sum_{j,l=1}^{m}\alpha_j^i\alpha_l^i\cdot k\left(\mathbf{x}_j,\mathbf{x}_l\right)\\&=\left(\boldsymbol{\alpha}^i\right)^t\mathbf{K}\boldsymbol{\alpha}^i\\&=\lambda^i\left(\boldsymbol{\alpha}^i\right)^t\boldsymbol{\alpha}^i\\&=\lambda^i\left\|\boldsymbol{\alpha}^i\right\|^2\\&=1\end{aligned}$$

因此只须使 $\left\|\boldsymbol{\alpha}^i\right\|^2=\dfrac{1}{\lambda^i}$，即可使 $\mathbf{V}_i$ 规范化。

计算样本 $\mathbf{x}$ 在特征空间中第 j 个轴上的投影，利用(11)式，有：

$$\phi^t\left(\mathbf{x}\right)\mathbf{V}_j=\sum_{i=1}^{m}\alpha_i^j\,\phi^t\left(\mathbf{x}\right)\phi\left(\mathbf{x}_i\right)=\sum_{i=1}^{m}\alpha_i^j\,k\left(\mathbf{x},\mathbf{x}_i\right)\tag{12}$$

KPCA 算法：

1. 计算矩阵；$\mathbf{K}=\left[k\left(\mathbf{x}_i,\mathbf{x}_j\right)\right]_{m\times m}$；
2. 计算 $\mathbf{K}$ 的特征值和特征向量：$\lambda^1,\cdots,\lambda^m$，$\boldsymbol{\alpha}^1,\cdots,\boldsymbol{\alpha}^m$，取最大的前 p 个特征值 $\lambda^1,\lambda^2,\ \ldots\ ,\lambda^p$；
3. 规范化前 p 个特征向量，使得 $\left\|\boldsymbol{\alpha}^i\right\|^2=\dfrac{1}{\lambda^i}$；
4. 利用(12)式可以计算样本 $\mathbf{x}$ 的第 j 个核主成分。