- 点到点之间的距离

在 n 维空间中，a 与 b 两点之间的欧氏距离为：

$$D(a, b) = \| a - b \|$$

写成距离平方：

$$D^2(a,b) = (a-b)^T(a-b) = \sum_{k=1}^{n}(a_k - b_k)^2$$

其中，a 和 b 为 n 维向量，其第 k 个分量分别是 a_k 和 b_k。

- 点到点集之间的距离

在 n 维空间中，点 x 到点 $a^{(i)}$ 之间的距离平方为：

$$D^2(x, a^{(i)}) = \sum_{k=1}^{n}(x_k - a_k^{(i)})^2$$

因此，点 x 到点集 $\{a^{(i)}\}_{i=1,2,\cdots,K}$ 之间的均方距离为：

$$\overline{D^2(x, \{a^{(i)}\})} = \frac{1}{K}\sum_{i=1}^{K} D^2(x, a^{(i)}) = \frac{1}{K}\sum_{i=1}^{K}\left\{\sum_{k=1}^{n}(x_k - a_k^{(i)})^2\right\}$$

- 类内距离

n 维空间中同一类内各模式样本点集$\{a^{(i)}\}_{i=1,2,\cdots,K}$，其内部各点的均方距离为$\overline{D^2(\{a^{(j)}\},\{a^{(i)}\})}$，其中$i,j=1,2,\ldots,K,i\neq j$，即：

$$\overline{D^2(\{a^{(j)}\},\{a^{(i)}\})}=\frac{1}{K}\sum_{j=1}^{K}\left[\frac{1}{K-1}\sum_{\substack{i=1\\i\neq j}}^{K}\sum_{k=1}^{n}(a_k^{(j)}-a_k^{(i)})^2\right]$$

可证明：

$$\overline{D^2}=2\sum_{k=1}^{n}\sigma_k^2$$

其中σ_k^2为$\{a^{(i)}\}$在第 k 个分量上的无偏方差，即：

$$\sigma_k^2=\frac{1}{K-1}\sum_{i=1}^{K}(a_k^{(i)}-\overline{a_k})^2$$

其中$\overline{a_k}=\dfrac{1}{K}\sum_{i=1}^{K}a_k^{(i)}$为$\{a^{(i)}\}$在第 k 个分量方向上的均值。

[证明作为练习]

- 类内散布矩阵

考虑一类内模式点集$\{a^{(i)}\}_{i=1,2,\dots,K}$，其类内散布矩阵为：

$$S=\sum_{i=1}^{K}\{(a^{(i)}-m)(a^{(i)}-m)^{T}\}$$

其中

$$m=\frac{1}{K}\sum_{i=1}^{K}a^{(i)}$$

对属于同一类的模式样本，类内散布矩阵表示各样本点围绕其均值周围的散布情况。

● 类间距离和类间散布矩阵

在考虑有两个以上的类别，如集合$\{a^{(i)}\}$和$\{b^{(j)}\}$时，类间距离对类别的可分性起着重要作用，此时应计算：

$$\overline{D^2(\{a^{(i)}\},\{b^{(j)}\})}_{i=1,2,\dots,K_a;j=1,2,\dots,K_b}$$

为简化起见，常用两类样本各自质心间的距离作为类间距离，并假设两类样本出现的概率相等，则：

$$D^2=\sum_{k=1}^{n}(m_{1_k}-m_{2_k})^2$$

其中m_1和m_2为两类模式样本集各自的均值向量，m_{1_k}和m_{2_k}为m_1和m_2的第 k 个分量，n 为维数。

写成矩阵形式：$S_{b2}=(m_1-m_2)(m_1-m_2)^T$为两类模式的类间散布矩阵。

对三个以上的类别，类间散布矩阵常写成：

$$S_b=\sum_{i=1}^{c}P(\omega_i)(m_i-m_0)(m_i-m_0)^T$$

其中，m_0为多类模式（如共有 c 类）分布的总体均值向量，即：

$$m_0=E\{x\}=\sum_{i=1}^{c}P(\omega_i)m_i,\quad \forall\omega_i,\ i=1,2,\dots,c$$

- 多类模式集散布矩阵

多类情况的类内散布矩阵，可写成各类的类内散布矩阵的先验概率的加权和，即：

$$S_w = \sum_{i=1}^{c} P(\omega_i) E\{(x - m_i)(x - m_i)^T \mid \omega_i\} = \sum_{i=1}^{c} P(\omega_i) C_i$$

其中 C_i 是第 i 类的协方差矩阵。

有时，用多类模式总体分布的散布矩阵来反映其可分性，即：

$$S_t = E\{(x - m_0)(x - m_0)^T\}, \quad x \in \forall\omega_i,\ i = 1,2,\ldots,c$$

其中，m_0 为多类模式分布的总体均值向量。

可以证明：$S_t = S_w + S_b$，即总体散布矩阵是各类类内散布矩阵与类间散布矩阵之和。

- 对于ω_i和ω_j两类训练样本的特征选择

例：对于ω_i和ω_j两类训练样本，假设其均值向量为m_i和m_j，其k维方向的分量为m_{ik}和m_{jk}，方差为σ_{ik}^2和σ_{jk}^2，定义可分性准则函数：

$$G_k = \frac{(m_{ik}-m_{jk})^2}{\sigma_{ik}^2+\sigma_{jk}^2},\quad k = 1,2,\dots,n$$

则G_K为正值。G_K值越大，表示测度值的第k个分量对分离ω_i和ω_j两类越有效。将$\{G_K, k=1,2,\dots,n\}$按大小排队，选出最大的m个对应的测度值作为分类特征，即达到特征选择的目的。

- 一般特征的散布矩阵准则

类内：$S_w = \sum_{i=1}^{c} P(\omega_i) E\{(x - m_i)(x - m_i)^T \mid \omega_i\}$

类间：$S_b = \sum_{i=1}^{c} P(\omega_i)(m_i - m_0)(m_i - m_0)^T$

直观上，类间离散度越大且类内离散度越小，则可分性越好。因此，可推导出散布矩阵准则采用如下形式：

行列式形式：$J_1 = \det(S_w^{-1} S_b) = \prod_i \lambda_i$

迹形式：$J_2 = tr(S_w^{-1} S_b) = \sum_i \lambda_i$

其中，λ_i 是矩阵 $S_w^{-1} S_b$ 的特征值。使 J_1 或 J_2 最大的子集可作为选择的分类特征。

- 离散的有限 K-L 展开式的形式

设一连续的随机实函数 x(t)， $T_1 \le t \le T_2$，则 x(t)可用已知的正交函数集{$\phi_j(t)$， j=1, 2, …}的线性组合来展开，即：

$$x(t) = a_1\varphi_1(t) + a_2\varphi_2(t) + \cdots + a_j\varphi_j(t) + \cdots$$

$$= \sum_{j=1}^{\infty} a_j\varphi_j(t), \quad T_1 \le t \le T_2 \qquad (1)$$

式中，a_j为展开式的随机系数，$\phi_j(t)$为一连续的正交函数，它应满足：

$$\int_{T_1}^{T_2} \varphi_n(t)\tilde{\varphi}_m(t)dt = \begin{cases} 1, & if\ m = n \\ 0 & if\ m \ne n \end{cases}$$

其中$\tilde{\phi}_m(t)$为$\phi_m(t)$的共轭复数式。

将上式写成离散的正交函数形式，使连续随机函数 x(t)和连续正交函数$\phi_j(t)$在区间$T_1 \le t \le T_2$内被等间隔采样为 n 个离散点，即：

$$x(t) \to \{x(1), x(2), \cdots, x(n)\}$$

$$\varphi_j(t) \to \{\varphi_j(1), \varphi_j(2), \cdots, \varphi_j(n)\}$$

写成向量形式：

$$x = (x(1), x(2), \cdots, x(n))^T$$

$$\varphi_j = (\varphi_j(1), \varphi_j(2), \cdots, \varphi_j(n))^T,\ j = 1,2,\cdots,n$$

将式(1)取 n 项近似，并写成离散展开式：

$$x = \sum_{j=1}^{n} a_j\phi_j = \Phi a, \quad T_1 \le t \le T_2 \qquad (2)$$

其中，a 为展开式中随机系数的向量形式，即：

$$a = (a_1, a_2, \ldots, a_j, \ldots, a_n)^T$$

Φ为 n x n 维矩阵，即：

$$\Phi = (\varphi_1, \varphi_2, \cdots, \varphi_n) = \begin{bmatrix} \varphi_1(1) & \varphi_2(1) & \cdots & \varphi_n(1) \\ \varphi_1(2) & \varphi_2(2) & \cdots & \varphi_n(2) \\ \cdots & \cdots & \cdots & \cdots \\ \varphi_1(n) & \varphi_2(n) & \cdots & \varphi_n(n) \end{bmatrix}$$

其中，每一列为正交函数集中的一个函数，小括号内的序号为正交函数的采样点次序。因此，Φ实质上是由 ϕ_j 向量组成的正交变换矩阵，它将 x 变换成 a。

● 正交向量集$\{\Phi_j\}$的确定

设随机向量 x 的总体自相关矩阵为$R=E\{xx^T\}$。由

$$x=\sum_{j=1}^{n}a_j\phi_j=\Phi a,\quad T_1\le t\le T_2 \tag{1}$$

将$x=\Phi a$代入$R=E\{xx^T\}$，得：

$$R=E\{\Phi a\, a^T\Phi^T\}=\Phi(E\{a\, a^T\})\Phi^T$$

要求系数向量 a 的各个不同分量应统计独立，即应使$(a_1, a_2, \ldots, a_j, \ldots, a_n)$满足如下关系：

$$E(a_j a_k)=\begin{cases}\lambda_j & \text{if } j=k\\ 0 & \text{if } j\ne k\end{cases}$$

写成矩阵形式，应使：$E\{a\, a^T\}=D_\lambda$，其中D_λ为对角形矩阵，其互相关成分均为 0，即：

$$D_\lambda=\begin{bmatrix}\lambda_1 & 0 & \cdots & \cdots & 0\\ 0 & \ddots & 0 & \cdots & 0\\ 0 & & \lambda_j & & 0\\ 0 & \cdots & 0 & \ddots & 0\\ 0 & \cdots & \cdots & 0 & \lambda_n\end{bmatrix}$$

则：

$$R=\Phi D_\lambda\Phi^T$$

由于Φ中的各个向量Φ_j都相互归一正交，故有：

$$R\Phi=\Phi D_\lambda\Phi^T\Phi=\Phi D_\lambda$$

其中，Φ_j向量对应为：

$$R\phi_j=\lambda_j\phi_j$$

可以看出，λ_j 是 x 的自相关矩阵 R 的特征值，ϕ_j 是对应的特征向量。因为 R 是实对称矩阵，其不同特征值对应的特征向量应正交，即：

$$\phi_j^T\phi_k=\begin{cases}1 & \text{if } j=k\\ 0 & \text{if } j\neq k\end{cases}$$

由式(1)，K-L 展开式系数应为：

$$a=\Phi^T x$$

- K-L 展开式系数的计算步骤

K-L 展开式系数可如下求得：

1. 求随机向量 x 的自相关矩阵：$R = E\{xx^T\}$
2. 求出矩阵 R 的特征值 λ_j 和对应的特征向量 ϕ_j，$j = 1,2,\ldots,n$，得矩阵：

$$\Phi = (\varphi_1, \varphi_2, \cdots, \varphi_n)$$

3. 计算展开式系数：

$$a = \Phi^T x$$

- 问题：选取变换矩阵Φ，使得降维后的新向量在最小均方差条件下接近原来的向量 x

对于 $x=\sum_{j=1}^{n}a_j\phi_j$，现仅取 m 项，对略去的系数项用预先选定的常数 b 代替，此时对 x 的估计值为：

$$\hat{x}=\sum_{j=1}^{m}a_j\phi_j+\sum_{j=m+1}^{n}b\phi_j$$

则产生的误差为：

$$\Delta x=x-\hat{x}=\sum_{j=m+1}^{n}(a_j-b)\phi_j$$

则Δx 的均方误差为：

$$\overline{\varepsilon^2}=E\{\|\Delta x\|\}^2=\sum_{j=m+1}^{n}\{E(a_j-b)^2\}$$

要使 $\overline{\varepsilon^2}$ 最小，对 b 的选择应满足：

$$\frac{\partial}{\partial b}[E(a_j-b)^2]=\frac{\partial}{\partial b}[E(a_j^2-2a_jb+b^2)]=-2[E(a_j)-b]=0$$

因此，$b=E[a_j]$，即对省略掉的 a 中的分量，应使用它们的数学期望来代替，此时的误差为：

$$\overline{\varepsilon^2}=\sum_{j=m+1}^{n}E[(a_j-E\{a_j\})^2]=\sum_{j=m+1}^{n}\phi_j^T E[(x-E\{x\})(x-E\{x\})^T]\phi_j$$

$$=\sum_{j=m+1}^{n}\phi_j^T C_x\phi_j$$

其中，C_x 为 x 的协方差矩阵

设 λ_j 为 C_x 的第 j 个特征值，ϕ_j 是与 λ_j 对应的特征向量，则

$$C_x\phi_j=\lambda_j\phi_j$$

由于

$$\phi_j^T \phi_j = 1$$

从而

$$\phi_j^T C_x \phi_j = \lambda_j$$

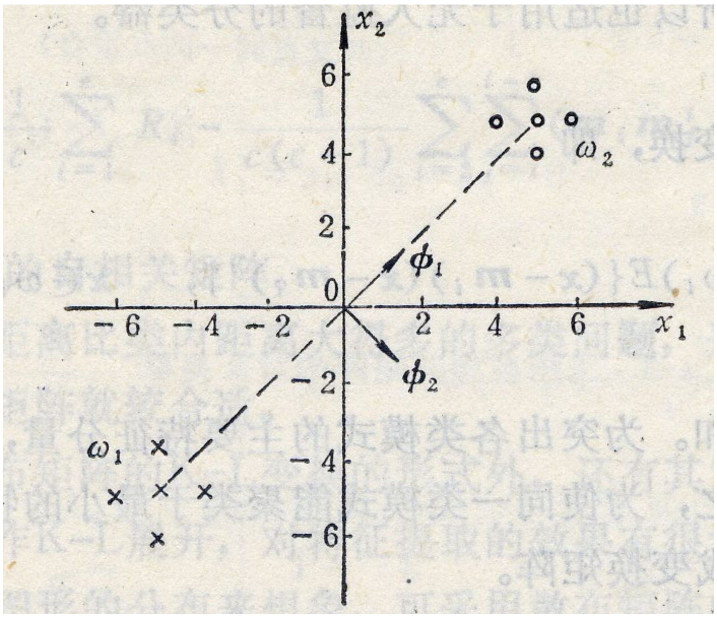
因此

$$\overline{\varepsilon^2} = \sum_{j=m+1}^{n} \phi_j^T C_x \phi_j = \sum_{j=m+1}^{n} \lambda_j$$

由此可以看出，λ_j值越小，误差也越小。

- K-L 变换实例

给定两类模式，其分布如图所示，试用 K-L 变换实现一维的特征提取（假定两类模式出现的概率相等）。



$$m=\frac{1}{5}\sum_{j=1}^{N_i}x_{1j}+\frac{1}{5}\sum_{j=1}^{N_i}x_{2j}=0$$

符合 K-L 变换进行特征压缩的最佳条件。

因 $P(\omega_1)=P(\omega_2)=0.5$，故

$$R=\sum_{i=1}^{2}P(\omega_i)E\{xx^T\}=\frac{1}{2}\left[\frac{1}{5}\sum_{j=1}^{5}x_{1j}x_{1j}^T\right]+\frac{1}{2}\left[\frac{1}{5}\sum_{j=1}^{5}x_{2j}x_{2j}^T\right]=\begin{pmatrix}25.4 & 25.0\\ 25.0 & 25.4\end{pmatrix}$$

解特征值方程 $|R-\lambda I|=0$，求 R 的特征值。

由 $(25.4-\lambda)^2-25.0^2=0$，得特征值 $\lambda_1=50.4$，$\lambda_2=0.4$

其对应的特征向量可由 $R\Phi_i=\lambda_i\Phi_i$ 求得：

$$\Phi_1=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix},\quad \Phi_2=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$$

选 λ_1 对应的变换向量作为变换矩阵，由 $y=\Phi^T x$ 得变换后的一维模式特征为：

$$\omega_1 : \left\{ -\frac{10}{\sqrt{2}}, -\frac{9}{\sqrt{2}}, -\frac{9}{\sqrt{2}}, -\frac{11}{\sqrt{2}}, -\frac{11}{\sqrt{2}} \right\} \; \omega_2 : \left\{ \frac{10}{\sqrt{2}}, \frac{11}{\sqrt{2}}, \frac{11}{\sqrt{2}}, \frac{9}{\sqrt{2}}, \frac{9}{\sqrt{2}} \right\}$$