一、 名词解释：

1. 特征选择与特征提取

名词约定：

- 特征形成（特征获取、提取）

  直接观测到的或经过初步运算的特征—

- 特征选择

  从 m 个特征中选择 m_1 个，$m_1 < m$ （人

- 特征提取（特征变换，特征压缩）

  将 m 个特征变为 m_2 个新特征 -

2. 特征的评价准则

   概念：数学上定义的用以衡量特征对分类的效果的准则，实际问题中需根据实际情况人为确定 。'

   误识率判据：理论上的目标，实际采用困难（密度未知，形式复杂，样本不充分，…）

   可分性判据：实用的可计算的判据

3. 总类内离散度矩阵S_w,总类间离散度矩阵S_b。（写出计算公式）

类间平均距离：

$$J_D = \frac{1}{2}\sum_{i=1}^{c} P_i \sum_{j=1}^{c} P_j \frac{1}{n_i n_j}\sum_{k=1}^{n_i} \sum_{l=1}^{n_j} \delta\left(x_k^{(\cdot}\right.$$

其中， $x_k^{(i)} \in \omega_i$， k

$x_l^{(j)} \in$

P_i， P_j

$\delta(x_k, x$

度量

通常采用欧氏距离：

$$\delta(x_k, x_l) = (x_k - x_l)^T (x_k -$$

J_D 称为各类之间的平均平方距离

定义：　　类均值向量　　$m_i = \frac{1}{n_i} \sum_{k=1}^{n_i}$

总均值向量　　$m = \sum_{i}$

类间离散度矩阵 S_b 的估计：

$$\widetilde{S}_b = \sum_{i=1}^{c} P_i (m_i - m)(m_i - m)^T$$

类内离散度矩阵 S_w 的估计：

$$\widetilde{S}_w = \sum_{i=1}^{c} P_i \frac{1}{n_i} \sum_{k=1}^{n_i} \left(x_k^{(i)} - m_i\right)\left(x_k^{(i)} - m_i\right)^T = \sum_{i}$$

Σ_i：类协方

则　　$J_D = \mathrm{tr}\left(\widetilde{\mathrm{S}}_{\mathrm{w}} + \widetilde{S}_b\right)$

4. 基于类内类间距离的可分性判据

$$\widetilde{S}_w = \sum_{i=1}^{c} P_i \frac{1}{n_i} \sum_{k=1}^{n_i} \left(x_k^{(i)} - m_i\right)\left(x_k^{(i)} - m_i\right)^T = \sum_{i}$$

Σ_i：类协方

则 $$J_D = \mathrm{tr}\left(\widetilde{\mathrm{S}}_{\mathrm{w}} + \widetilde{S}_b\right)$$

常用的基于类内类间距离的可分性判据：

$$J_1 = \mathrm{tr}\left(\mathrm{S}_{\mathrm{w}} + S_b\right)$$

$$J_2 = \mathrm{tr}\left(S_w^{-1} S_b\right)$$

$$J_3 = \ln \frac{\left|S_b\right|}{\left|S_w\right|}$$

5. 基于概率分布的可分性判据，基于熵的可分性判据

$$8J_B = J_D = (\mu_1 - \mu_2)^T \sum^{-1} (\mu_1 -$$

基于概率相关性的可分性判据

考查联合分布密度 $p(x, \omega_i) = p(x \mid \omega_1) P(\omega_i)$

如 x 与 ω_i 独立，$p(x, \omega_i) = p(x) p(\omega_i)$，即 $p(x) =$

则 x 不能提供对分类 ω_i 的信息

因此可定义 $p(x)$ 与 $p(x \mid \omega_i)$ 之间关系的一个函数作

$$J_i = \int g\left(p(x \mid \omega_i), p(x), P(\omega_i\right.$$

称作概率

7.2.3　基于熵的可分性判据

熵：事件不确定性的度量

A 事件的不确定性大（熵大），则对 A 事件

思路：

把各类 ω_i 看作一系列事件

把后验概率 $P(\omega_i \mid x)$ 看作特征 x 上出现 ω_i 的

如从 x 能确定 ω_i，则对 ω_i 的观察不提供信

—— 特征 x

如从 x 完全不能确定 ω_i，则对 ω_i 的观察信

—— 特征 x

定义熵函数 $H = J_c\left[P(\omega_1 \mid x), \cdots P(\omega_c \mid x)\right]$

须满足

①规一化 $J_c\left(\frac{1}{c}, \cdots, \frac{1}{c}\right) = 1$

$$0 \le J_c\left(P_1, \cdots, P_c\right) \le J_c\left(\frac{1}{c}, \cdots, \frac{1}{c}\right) = 1$$

②对称性 $J_c\left(P_1, \cdots, P_c\right) = J_c\left(P_?, \cdots, P_?\right)$

③确定性 $J_c(1,0,\cdots,0) = J_c(0,1,\cdots,0) = \cdots = 0$

④扩张性 $J_c\left(P_1, \cdots, P_c\right) = J_{c+1}\left(P_1, \cdots, P_c, 0\right)$

⑤连续性 $P(\omega_i \mid x)$ 的连续函数

⑥分枝性（综合性） 一分为二，则熵增加；二合为

Shannon 熵：

$$H = -\sum_{i=1}^{c} P(\omega_i \mid x)\log_2 P(\omega$$

平方熵：

$$H = 2\left[1-\sum_{i=1}^{c} P^2(\omega_i \mid x)\right]$$

熵可分离性判据： $$J_e = \int H(x)p(x)dx$$

J_e 大，则重叠性大，可分性不好，

J_e 小，则可分性好。

6. 主成分分析（简答）

   目的 出发点是从一组特征中计算出一组按重要性从大到小排列的新特征，他们是原有特征的线性组合，并且互相之间是不相关的。

7. K-L 变换的基本原理

   函数的级数展开：将函数用一组（正交）基函数展开，用展开系数表示原函数。

   离散 K－L 展开：　　把随机向量用一组正交基向量展开，用展开系数代表原向量。

   基向量所张成的空间：新的特征空间。

   展开系数组成的向量：新特征空间中的样本向量

二、　计算题

设有两类问题，其先验概率相等，即$P(\omega_1)$= $P(\omega_2)$=$\frac{1}{2}$，样本均值向量分别为$\mu_1 = [4,2]^T, \mu_2 = [-4,-2]^T$,协方差矩阵分别是$\Sigma_1 = \begin{bmatrix}3 & 1\\ 1 & 3\end{bmatrix}$ 和$\Sigma_2 = \begin{bmatrix}4 & 2\\ 2 & 4\end{bmatrix}$。试利用 K-L 变换把维数从 2 压缩为 1。

1 总体自相关矩阵 R

$$R=E\{\mathrm{XX}^T\}=\frac{1}{2}(\mu_1\mu_1^T+\mu_2\mu_2^T)=\begin{bmatrix}16 & 8\\ 8 & 4\end{bmatrix}$$

2 计算 R 的本征值 并选择较大者。 由$\left|R-\lambda I\right|=0$ 得

λ_1=24，λ_2=0

3. 根据 $R\mu_1=\lambda_1\mu_1$ 计算 λ_1 对应的特征向量 μ_1 ， 归一化后为

$$U=\left[\mathrm{u}\right]=\begin{bmatrix}0.7071\\ 0.7071\end{bmatrix}$$

4.利用 U 对样本集中的每一个样本进行 K-L 变换

$$X_1^*=U^T X_1=\begin{bmatrix}0.7071 & 0.7071\end{bmatrix}\begin{bmatrix}4\\ 2\end{bmatrix}=\begin{bmatrix}4.2426\end{bmatrix}$$

$$X_2^*=U^T X_2=\begin{bmatrix}0.7071 & 0.7071\end{bmatrix}\begin{bmatrix}-4\\ -2\end{bmatrix}=\begin{bmatrix}-4.2426\end{bmatrix}$$