

1、在图像识别中，假定有灌木和坦克 2 种类型，它们的先验概率分别是 0.7 和 0.3，损失函数如下表所示。其中，类型 w_1 和 w_2 分别表示灌木和坦克，判决 $a_1=w_1$ ， $a_2=w_2$ 。现在做了 2 次实验，获得 2 个样本的类概率密度如下：

$$P(x | \omega_1) = 0.2 \quad 0.5$$

$$P(x | \omega_2) = 0.6 \quad 0.3$$

损失 决策 \ 状态	w_1	w_2
a_1	0.5	2
a_2	4	1.0

(1) 试用最小错误率贝叶斯准则判决 2 个样本各属于哪一类？坦克、灌木。

(2) 试用最小风险决策规则判决 2 个样本各属于哪一类？灌木、灌木。

答：(1) 最小错误率贝叶斯准则

第一个样本：

$$P(\omega_1 | x) = \frac{p(x | \omega_1)P(\omega_1)}{\sum_{j=1}^2 p(x | \omega_j)P(\omega_j)} = \frac{0.2 * 0.7}{0.2 * 0.7 + 0.6 * 0.3} = \frac{14}{32} = 0.4375$$

$$P(\omega_2 | x) = 1 - P(\omega_1 | x) = 1 - 0.4375 = 0.5625$$

$$P(\omega_2 | x) > P(\omega_1 | x) \Rightarrow x \in \omega_2, \text{ 决策为坦克}$$

第二个样本：

$$P(\omega_1 | x) = \frac{p(x | \omega_1)P(\omega_1)}{\sum_{j=1}^2 p(x | \omega_j)P(\omega_j)} = \frac{0.5 * 0.7}{0.5 * 0.7 + 0.3 * 0.3} = \frac{35}{44} \approx 0.795$$

$$P(\omega_2 | x) = 1 - P(\omega_1 | x) \approx 1 - 0.795 = 0.205 = \frac{9}{44}$$

$$P(\omega_2 | x) < P(\omega_1 | x) \Rightarrow x \in \omega_1, \text{ 决策为灌木}$$

(2) 最小风险决策规则

$$\lambda_{11} = 0.5 \quad \lambda_{12} = 2 \quad \lambda_{21} = 4 \quad \lambda_{22} = 1.0$$

第一个样本

$$\begin{aligned} R(a_1 | x) &= \sum_{j=1}^2 \lambda_{1j} P(\omega_j | x) = \lambda_{11} P(\omega_1 | x) + \lambda_{12} P(\omega_2 | x) \\ &= 0.5 * 0.4375 + 2 * 0.5625 = 1.35375 \end{aligned}$$

$$\begin{aligned} R(a_2 | x) &= \sum_{j=1}^2 \lambda_{2j} P(\omega_j | x) = \lambda_{21} P(\omega_1 | x) + \lambda_{22} P(\omega_2 | x) \\ &= 4 * 0.4375 + 1.0 * 0.5625 = 2.3175 \end{aligned}$$

$$R(a_1 | x) < R(a_2 | x) \Rightarrow x \in \omega_1, \text{ 决策为灌木}$$

第二个样本

$$\begin{aligned} R(a_1 | x) &= \sum_{j=1}^2 \lambda_{1j} P(\omega_j | x) = \lambda_{11} P(\omega_1 | x) + \lambda_{12} P(\omega_2 | x) \\ &= 0.5 * 0.795 + 2 * 0.205 = 0.8075 \end{aligned}$$

$$\begin{aligned} R(a_2 | x) &= \sum_{j=1}^2 \lambda_{2j} P(\omega_j | x) = \lambda_{21} P(\omega_1 | x) + \lambda_{22} P(\omega_2 | x) \\ &= 4 * 0.795 + 1.0 * 0.205 = 3.385 \end{aligned}$$

$R(a_1 | x) < R(a_2 | x) \Rightarrow x \in \omega_1$, 决策为灌木

2、给出二维样本数据 $(-1,1),(2,2),(1,-1),(-2,-2)$, 试用 K-L 变换作一维数据压缩。

答：数据压缩结果： $0, 2\sqrt{2}, 0, -2\sqrt{2}$

1.样本的均值向量为：

$$m = \frac{1}{4} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

2.自相关矩阵

$$\begin{aligned} R &= \frac{1}{4} \left[\begin{pmatrix} -1 \\ 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 2 \end{pmatrix} \begin{pmatrix} 2 & 2 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} + \begin{pmatrix} -2 \\ -2 \end{pmatrix} \begin{pmatrix} -2 & -2 \end{pmatrix} \right] \\ &= \frac{1}{4} \begin{pmatrix} 10 & 6 \\ 6 & 10 \end{pmatrix} = \begin{pmatrix} 2.5 & 1.5 \\ 1.5 & 2.5 \end{pmatrix} \end{aligned}$$

3.求特征值与特征向量

$$\begin{vmatrix} \lambda - 2.5 & -1.5 \\ -1.5 & \lambda - 2.5 \end{vmatrix} = 0 \Rightarrow \lambda_1 = 4, \lambda_2 = 1$$

特征向量(标准)分别是：

$$X_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{2}{\sqrt{2}} \\ \frac{\sqrt{2}}{2} \\ \frac{2}{2} \end{pmatrix}, X_2 = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{2}{\sqrt{2}} \\ \frac{\sqrt{2}}{2} \\ \frac{2}{2} \end{pmatrix}$$

4.取更大的特征值所对应的特征向量 X_1 为变换矩阵

5.将原样本变换成一维样本(分别用 X_1^T 左乘以每原数据样本),得

$$\begin{pmatrix} -\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} * 2 + \frac{\sqrt{2}}{2} * 2 & \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} * (-2) - \frac{\sqrt{2}}{2} * (-2) \end{pmatrix} \\ = (0 \quad 2\sqrt{2} \quad 0 \quad -2\sqrt{2})$$

0---算出 m 后应该把它当作坐标原点重新计算其他坐标值

3--- $|\lambda E - A| = 0$

$(\lambda E - A) * X = 0$ 0 向量

- 平移坐标系, 将模式的总体均值向量作为新坐标系的原点
- 求随机向量 X 的自相关矩阵

- 求自相关矩阵的 n 个特征值及其对应的特征向量
- 将特征值从大到小排序, 取前 m 个大的特征值所对应的特征向量构成新的变换矩阵
- 将 n 维向量变换为 m 维新向量

3、已知两类数据: $\omega_1: (1,0),(2,0),(1,1)$; $\omega_2: (-1,0),(0,1),(-1,1)$, 试求该组数据的类内与类间散布矩阵。

答:

1).取均值向量

$$m_i = \frac{1}{N_i} \sum_{x \in \mathcal{X}_i} x$$

$$m_1 = \begin{pmatrix} \frac{4}{3} & \frac{1}{3} \end{pmatrix}^T, m_2 = \begin{pmatrix} -\frac{2}{3} & \frac{2}{3} \end{pmatrix}^T$$

2).分别计算两个类与均值向量的距离平方和

$$S_1 = \sum_{x \in \mathcal{X}_1} (x - m_1)(x - m_1)^T = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

$$\begin{bmatrix} -\frac{1}{3} & -\frac{1}{3} \end{bmatrix}^T \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \end{bmatrix}^T \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} \end{bmatrix}^T$$

$$S_2 = \sum_{x \in \mathcal{X}_2} (x - m_2)(x - m_2)^T = \frac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$\begin{bmatrix} -\frac{1}{3} & -\frac{2}{3} \end{bmatrix}^T \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}^T \begin{bmatrix} -\frac{1}{3} & \frac{1}{3} \end{bmatrix}^T$$

3).计算 S_w 与 S_b

$$S_w = S_1 + S_2 = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} + \frac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$$

$$S_b = (m_1 - m_2)(m_1 - m_2)^T = \begin{pmatrix} \frac{6}{3} \\ \frac{1}{3} \\ -\frac{1}{3} \end{pmatrix} \begin{pmatrix} \frac{6}{3} & -\frac{1}{3} \end{pmatrix} = \frac{1}{9} \begin{pmatrix} 36 & -6 \\ -6 & 1 \end{pmatrix}$$

S_w (within) neilei S_b (betwwen)neijian

4、已知欧氏二维空间中两类 9 个训练样本 $w_1: (-1,0)^T, (-2,0)^T, (-2,1)^T, (-2,-1)^T$

$w_2: (1,1)^T, (2,0)^T, (1,-1)^T, (2,1)^T, (2,2)^T$, 试分别用最近邻法和 K 近邻法求测试样本 $(0,0)^T$ 的分类, 取 $K=5, 7$ 。

答:

最近邻法: 最近邻为 $(-1, 0)^T$ 分类为 w_1

K 近邻法:

$K=5$: 5 个近邻为 1 类的 $(-1,0)^T, (-2,0)^T$, 2 类的 $(1,1)^T, (2,0)^T, (1,-1)^T$ 分类为 w_2

$K=7$: 1) 若近邻为 1 类的 $(-1,0)^T, (-2,0)^T, (-2,1)^T, (-2,-1)^T$, 2 类的 $(1,1)^T, (2,0)^T, (1,-1)^T$, 则分类为 w_1

2) 若近邻为 1 类的 $(-1,0)^T, (-2,0)^T, (-2,1)^T$ 或 $(-2,-1)^T$ 两个之一, 2 类的 $(1,1)^T, (2,0)^T, (1,-1)^T, (2,1)^T$, 则分类为 w_2

2、已知两类的训练样本: $w_1(0,0)^T, (0,2)^T; w_2(2,0)^T, (2,2)^T$, 试用最小平方误差准则算法进行分类器训练, 求解向量 w^* 。

1. 什么是模式与模式识别【模式: 对象之间存在的规律性关系, 模式识别: 是研究用计算机来实现人类模式识别能力的一门学科】

2. 什么是误差平方和准则

对于一个给定的聚类, 均值向量是最能代表聚类中所有样本的一个向量, 也称其为聚类中心。一个好的聚类方法应能使集合中的所有向量与这个均值向量的误差的长度平方和最小。

3. 确定线性分类器的主要步骤

采集训练样本, 构成训练样本集。样本应该具有典型性

确定一个准则 $J=J(w,x)$, 能反映分类器性能, 且存在权值 w^* 使得分类器性能最优

设计求解 w 的最优算法, 得到解向量 w^*

4. 分级聚类算法的 2 种基本途径是什么

按事物的相似性, 或内在联系组织起来, 组成有层次的结构, 使得本质上最接近的划为一类, 然后把相近的类再合并, 依次类推, 这就是分级聚类算法的基本思想。

聚合法: 把所有样本各自看为一类, 逐级聚合成一类。基本思路是根据类间相似性大小逐级聚合, 每级只把相似性最大的两类聚合成一类, 最终把所有样本聚合为一类。

分解法: 把所有样本看做一类, 逐级分解为每个样本一类。

5. 什么是 K 近邻法

取未知样本 x 的 k 个近邻, 看这 k 个近邻中多数属于哪一类, 就把 x 归为哪一类。

6. 监督学习与非监督学习的区别

利用已经标定类别的样本集进行分类器设计的方法称为监督学习。很多情况下无法预先知道

样本的类别，从没有标记的样本集开始进行分类器设计，这就是非监督学习

7. 什么是支持向量机

过两类样本中离分类面最近的点且平行于最优分类面的超平面上的训练样本，叫做支持向量。

支持向量机的基本思想：首先通过非线性变换将输入空间变换到一个高维空间，然后在这个新空间中求取最优线性分类面，而这种非线性变换是通过定义适当的内积函数实现的。

8. 近邻法的基本思想是什么

作为一种分段线性判别函数的极端情况，将各类中全部样本都作为代表点，这样的决策方法就是近邻法的基本思想。

9. 描述 K 均值聚类算法

给定一个数据点集合和需要的聚类数目 k ， k 由用户指定， k 均值算法根据某个距离函数反复把数据分入 k 个聚类中。先随机选取 K 个对象作为初始的聚类中心。然后计算每个对象与各个种子聚类中心之间的距离，把每个对象分配给距离它最近的聚类中心。聚类中心以及分配给它们的对象就代表一个聚类。一旦全部对象都被分配了，每个聚类的聚类中心会根据聚类中现有的对象被重新计算。

10. 详细写出感知器训练算法步骤

给定初始值：置 $k=0$ ，权向量 $w(k)$ 为任意值，可选常数 $0 < c \leq 1$

输入样本 $x_m \in \{x_1, x_2, \dots, x_n\}$ ，计算判决函数值 $g(x_m) = w^T(k)x_m$

按如下规则修改权向量

若 $x_m \in w_i$ ，且 $g(x_m) \leq 0$ ，则 $w(k+1) = w(k) + cx_m$

若 $x_m \in w_j$ ，且 $g(x_m) > 0$ ，则 $w(k+1) = w(k) - cx_m$

令 $k=k+1$ ，返回第二步，直到 w 对所有样本稳定不变，结束

11. 详细介绍初始聚类中心的选择方法

任取前 c 个样本点作为初始聚类中心

凭经验选择

将全部数据随机分为 c 类，计算其重心，将重心作为聚类中心

密度法选择代表点（具有统计特性）

从 $c-1$ 类划分中产生 c 类划分问题的初始聚类中心

12. 详细写出 Fisher 算法步骤

$$\begin{aligned}
1. & \quad \text{把两类样本分为 } X_1 \text{ 与 } X_2 \\
2. & \quad M_i = \frac{1}{n} \sum_{x_k \in X_i} x_k, i = 1, 2 \\
3. & \quad S_i = \frac{1}{n} \sum_{x_k \in X_i} (x_k - M_i)(x_k - M_i)^T, i = 1, 2 \\
4. & \quad S_w = S_1 + S_2 \\
5. & \quad w^* = \frac{1}{S_w} (M_1 - M_2)
\end{aligned}$$

13. 什么是两分剪辑近邻法与压缩近邻法

将原始样本随机分为两个集合：预测集 **T** 和参考集 **R**，来自预测集和参考集的样本分别完成考试和参考任务，相互独立

在剪辑的基础上，再去掉一部分这样的样本，有助于进一步缩短计算时间和降低存储要求。这类方法叫作压缩近邻法。

=====

2、一个典型的模式识别系统主要由哪几个部分组成？

原始数据获取和预处理、特征提取与选择、分类或聚类、后处理 **4** 部分

4、什么是后验概率？

系统在某个具体的模式样本 **X** 条件下位于某种类型的概率。

5、样本集推断总体概率分布的方法？

1、参数估计

- i. 监督参数估计：样本所属类别及类条件总体概率密度函数的形式已知，某些参数未知
- ii. 非监督参数估计：已知总体概率密度函数形式但未知样本类别，要推断某些参数

非参数估计：已知样本类别，未知总体概率密度函数形式，要求直接推断概率密度函数本身

11、特征抽取与特征选择的区别？

原始特征的数量可能很大，或者样本处于一个高维空间中，通过映射（或变换）的方法可以用低维空间来表示样本，这个过程叫特征抽取。

特征选择：从一组特征中挑选出一些最有效的特征以达到降低特征空间维数的目的，这个过程叫特征选择。

12、什么是最优搜索算法？

至今能得到最优解的唯一快速算法是“分支定界”算法。属于自上而下的算法，具有回溯功能。

13、统计学习理论的核心问题？

经验风险最小化原则下统计学习一致性的条件

在这些条件下关于统计学习方法推广性的界的结论

在这些界的基础上建立的小样本归纳推理原则

实现这些新的原则的实际方法

1、描述贝叶斯公式及其主要作用。

贝叶斯公式：两个事物 X 与 w 联合出现的概率称为联合概率。利用该公式可以计算后验概率。

$$P(X, \omega) = P(X | \omega)P(\omega) = P(\omega | X)P(X)$$

2、利用最大似然估计方法对单变量正态分布函数来估计其均值 μ 和方差 σ^2 。

P48

8、什么是离散 K-L 变换以及离散有限 K-L 展开。

一种基于目标统计特性的最佳正交变换；

9、必考：针对某个识别对象设计自己的模式识别系统，并叙述各步骤主要工作。