

—武大本本科生课程



“MLPR第4讲、第5讲”**贝叶斯决策**
——**概率论基础知识回顾及补充**

武汉大学计算机学院机器学习课程组

一. 模式识别的两类研究对象

模式识别的重要目的是要确定某一给定的模式样本属于哪一类。

我们通过对被识别对象的多种观察和测量构成特征向量，并将其作为某一判决规则的输入，依此规则对样本进行分类。在这个过程中，最初获取模式观察值时会遇到两种情况：

1. 事物间有确定的因果关系-确定性事件

即某一事件在一定条件下必然会发生或必然不发生。如，根据“三条直线边的闭合和一个直角”这一特征，就完全可以确定是直角三角形，这是**确定性现象**(Deterministic phenomena)。“第3章 线性判别分析”的模式判别就是基于这类现象，一个模式要么属于这一类，要么属于其他类。

2.事物间没有确定的因果关系-随机事件

在许多实际情况中，由于存在噪声和缺乏测度模式向量的完整信息，有些观察数据具有不确定的特点，有时属于某一类，有时又不属于该类，只有在大量重复的观察下才会呈现出某种规律性。也就是说，对它们观察得到的特征具有统计特性，特征向量不再是一个确定的向量，而是随机向量，其分量是随机变量。

基于随机现象(Random/Stochastic phenomena)的随机模式向量只能利用模式集的统计特性来分类，以使分类器发生分类错误的概率最小，这就是“基于统计决策理论的概率分类法”所要讨论的问题。这时不能说一个模式一定属于某一类，只能说它属于某一类的可能性(概率)有多大。

二. 概率论基础(知识回顾)

1. 概率的性质

1) 不可能事件 \emptyset 的概率为0, 即 $P(\emptyset)=0$;

2) $P(\bar{A})=1-P(A)$, 这里 \bar{A} 为 A 的补事件或 A 的对立事件;

3) 设 A, B 是两个随机事件, 有 $P(A \cup B)=P(A)+P(B)-P(AB)$;

其中, $P(AB)$ 为 A, B 同时发生的概率/联合概率, $P(A \cup B)$ 为 A, B 的并事件的概率.

2. 条件概率

设 A, B 为两个随机事件, $P(B) > 0$, 称已知事件 B 发生条件下事件 A 发生的条件概率为 $P(A|B) = \frac{P(AB)}{P(B)}$.

3.条件概率的三个重要公式

1)概率乘法公式(Product rule, 乘积规则)

若 $P(A) > 0$, 则A, B的联合概率 $P(AB) = P(A)P(B|A)$;

若 $P(B) > 0$, 则A, B的联合概率 $P(AB) = P(B)P(A|B)$;

2) 全概率公式

设事件 A_1, A_2, \dots, A_n 两两互斥, 且

$$\sum_{i=1}^n A_i = \Omega, P(A_i) > 0, i = 1, 2, \dots, n$$

则对任一事件B, 有如下全概率公式:

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

直观上看, 全概率公式是从“原因” A_i 推出“结果”B发生的概率计算公式。

3) 贝叶斯公式 (*Bayes Theorem*)

在前面全概率公式的条件下, 若再有 $P(B) > 0$, 则有如下贝叶斯公式:

$$P(A_i|B) = \frac{P(A_i B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^n P(A_i)P(B|A_i)}$$

直观上, **贝叶斯公式**是已知“结果” B 发生, 推出某个“原因” A_i 发生的后验概率。

三.模式识别中常用的三个概率

1. 先验概率 $P(\omega_i)$

指事先根据大量的统计资料得出的 ω_i 类样本出现的概率.

先验概率来源于先前的知识和经验, 与现在无关, 提供的分类信息很少.

2. 后验概率 $P(\omega_i | \mathbf{x})$

后验概率与先验概率相对应, 指得到一批观察样本数据后统计出的 \mathbf{x} 属于 ω_i 类的概率.

3. 条件概率 $P(\mathbf{x} | \omega_i)$

指已知属于 ω_i 的样本 \mathbf{x} 发生某种事件的概率.

分类中常用条件概率密度/类条件概率密度, 也称条件概率密度函数/类概率密度函数, 统计学中称为似然函数(Likelihood function). ω_i 类的类概率密度函数表示为 $p(\mathbf{x} | \omega_i)$.

如,要对一批患者进行一项化验,可以用 ω_1 代表患病人群,患者的化验结果就是特征向量的值,仍用 \mathbf{x} 表示. 由于化验结果不是阴性就是阳性,因此这里的 \mathbf{x} 是一维特征向量,只有两个取值. 那么“对一批患者进行一项化验,结果为阳性的概率为95%”可以表示为 $P(\mathbf{x}=\text{阳性} \mid \omega_1)=0.95$;也可以先设 $\mathbf{x}=\text{阳性}$,写成 $P(\mathbf{x} \mid \omega_1)=0.95$.

分类中常用条件概率密度/类条件概率密度,也称条件概率密度函数/类概率密度函数,统计学中称为似然函数(Likelihood function). ω_i 类的类概率密度函数表示为 $p(\mathbf{x} \mid \omega_i)$.

例：一个二类问题， ω_1 类表示某地区患有某病的人群， ω_2 类表示无此病的人群.

那么：

先验概率 $P(\omega_1)$ 表示该地区居民患有此病的概率；

先验概率 $P(\omega_2)$ 表示该地区无此病的概率；

这两个值可以通过大量的统计资料得到.

如果采用某种方法检测是否患病，设 \mathbf{x} 表示“试验反应呈阳性”，那么：

$P(\mathbf{x}|\omega_2)$ 表示无病的人群做该实验时反应呈阳性(显示有病)的概率；

$P(\omega_2|\mathbf{x})$ 表示试验反应呈阳性的人中, 实际无病者的概率.

从上面的概率可以看出，诊断病情需要多种手段，用一种方法诊断为可能有病时，还要综合其他的结果才能最后确诊。

类似地，也有 ω_i 类的条件概率和后验概率。

4. 三个概率之间的关系

设有 M 类模式，根据贝叶斯定理 (*Bayes Theorem*)，可以得到后验概率，先验概率和类条件概率密度函数之间的关系为：

$$p(\omega_i | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_i) P(\omega_i)}{P(\mathbf{x})} = \frac{p(\mathbf{x} | \omega_i) P(\omega_i)}{\sum_{i=1}^M p(\mathbf{x} | \omega_i) P(\omega_i)}$$

$$1) p(\mathbf{x}) = \sum_{i=1}^M p(\mathbf{x}, \omega_i) = \sum_{i=1}^M p(\mathbf{x} | \omega_i) P(\omega_i) \text{ --- Sum rule (加和规则)}$$

这里 $p(\mathbf{x})$ 为边缘概率 (*marginal probability*)，简述为" \mathbf{x} 的概率"

$$2) p(\mathbf{x}, \omega_i) = p(\mathbf{x} | \omega_i) P(\omega_i) \text{ --- Product rule (乘积规则)}$$

四. 正态分布模式的贝叶斯决策数学基础

正态分布广泛存在于自然、生产及科学技术的众多领域中，对许多实际情况都是一种合适的模型。同时，正态分布又具有很多好的性质，有利于做数学分析，因此受到人们的高度重视。它在19世纪前叶由高斯加以推广，所以**又称为高斯分布**。

如果特征空间中的某一类样本较多地分布在其均值附近，远离均值点的样本较少，此时用正态分布作为概率模型是合理的。

前面的贝叶斯方法应用范围很广,但事先必须求出 $p(\mathbf{x} | \omega_i)$ 和 $P(\omega_i)$ 才能做出判决,这一工作一般做起来比较繁杂. 当 $p(\mathbf{x} | \omega_i)$ 呈现正态分布时,将会使决策简化,这时不再需要求 $p(\mathbf{x} | \omega_i)$ 的具体函数形式,只需要知道它的均值向量 $\boldsymbol{\mu}$ 和协方差矩阵 Σ 这两个参数即可.

相关知识介绍:

1. 二次型

设向量 $\mathbf{x}=[x_1, x_2, \dots, x_n]^T$, 矩阵 $\mathbf{A} = \begin{bmatrix} a_{11} & K & a_{1n} \\ M & 0 & M \\ a_{n1} & L & a_{nn} \end{bmatrix}$, 则 $\mathbf{x}^T \mathbf{A} \mathbf{x}$ 称为二次型.

它表示一个二次齐次多项式, 即 $\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i,j=1}^n a_{ij} x_i x_j$. 二次型中的矩阵 \mathbf{A} 是一个对称矩阵.

2. 正定二次型

$\forall \mathbf{x} \neq \mathbf{0}$ (\mathbf{x} 分量不全为零), 总有 $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$, 则称此二次型是正定的, 而对应的矩阵 \mathbf{A} 称为正定矩阵.

3. 单变量正态分布

单变量正态分布的概率密度函数定义为

$$\begin{aligned} p(x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \end{aligned}$$

式中, μ 为随机变量 x 的期望/均值; σ^2 为 x 的方差; σ 为标准差.

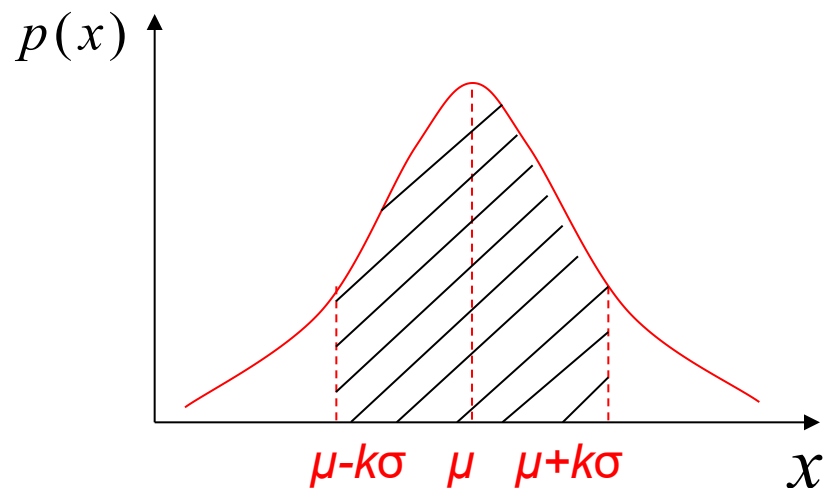
$$\mu = E(x) = \int_{-\infty}^{\infty} xp(x)dx$$

$$\sigma^2 = E\{(x-\mu)^2\} = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx$$

μ 一定时, 曲线的形状由 σ 确定. σ 越大, 曲线越“矮胖”, 表明总体分布越分散; 反之 σ 越小, 曲线越“瘦高”, 表明总体分布越集中.

4. 3σ 原则

$$P\{\mu - k\sigma \leq x \leq \mu + k\sigma\} = \begin{cases} 0.683 & k=1 \\ 0.954 & k=2 \\ 0.997 & k=3 \end{cases}$$



如右图所示,曲线下方的阴影部分的面积为概率 P 的值. 上式表明从正态总体中抽取的样本绝大部分都落在均值 μ 附近 $\pm 3\sigma$ 的范围内,因此正态分布概率密度曲线完全可以由均值和方差来确定,常简记为:

$$p(x) \sim N(\mu, \sigma^2)$$

服从正态分布 $N(\mu, \sigma^2)$ 的随机变量 x 取在 $\mu \pm 3\sigma$ 范围内的概率几乎达到1,这就是 3σ 原则. 3σ 原则比较定量地说明了正态分布的“两头小,中间大”的特点.

5. 多变量正态分布

多变量正态分布的概率密度函数与单变量类似, 定义为

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

式中: $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$, $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_n]^T$ 为均值向量;

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & K & \sigma_{1n}^2 \\ M & 0 & M \\ \sigma_{n1}^2 & L & \sigma_{nn}^2 \end{bmatrix} \text{ 为协方差矩阵, 是对称正定矩阵, 独立元素}$$

有 $n(n+1)/2$ 个. $|\Sigma|$ 为 Σ 的行列式, Σ^{-1} 为 Σ 的逆矩阵.

多维正态分布的概率密度函数完全由 n 个均值元素和协方差矩阵的 $n(n+1)/2$ 个独立元素所确定, 简记为 $p(\mathbf{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. 当 \mathbf{x} 的全部分量两两统计独立时, 协方差矩阵 $\boldsymbol{\Sigma}$ 全部非对角线上的元素都为零, 多变量正态分布概率密度函数可以简化成 n 个单变量正态分布概率函数的乘积.

以二维正态分布概率密度函数为例, 它的等密度线(等高线)投影到 x_1 和 x_2 面上为椭圆(见图所示), μ 是均值向量, 决定椭圆的位置. 椭圆的形状由协方差矩阵 Σ 决定, 椭圆在平行于 x_1 轴的方向上受 x_1 的方差 σ_{11}^2 的影响, 在平行于 x_2 轴的方向上受 x_2 的方差 σ_{22}^2 的影响, 在其他方向上受 x_1 和 x_2 的协方差 σ_{ij}^2 的影响, 这里 $i, j=1, 2$ 且 $i \neq j$. 椭圆的主轴方向由 Σ 的特征向量决定, 主轴的长度与相应的特征值成正比.

