模式识别的理论与方法

Pattern Recognition

裴继红

Chapter 2 (Part 1): Bayesian Decision Theory

(Sections 2.1-2.2)

概率、条件概率、联合概率

- 概率
- 条件概率
- 联合概率

Introduction

- 鲈鱼/鲑鱼的例子
  - 类别状态, 先验的(prior)

    类别状态是一个随机变量

    - 假设捕获的鲈鱼和鲑鱼数量是等概率的

      $P(\omega_1) = P(\omega_2)$ (等先验概率)

      $P(\omega_1) + P(\omega_2) = 1$ (排他性、穷尽性)

基本概念

- **类别状态.**

令 ω 表示类别状态. 则 ω 是一个随机变量.

(例如：对鲈鱼 $\omega = \omega_1$，对鲑鱼 $\omega = \omega_2$)

- **先验概率.**

令 $P(\omega_1)$ 和 $P(\omega_2)$ 分别表示 ω_1 和 ω_2 的先验概率.

简单决策规则

- **决策规则**

  在只有先验信息的条件下.

  若$P(\omega_1) > P(\omega_2)$，则归类为 ω_1; 负责归类为 ω_2

  但是，一般来说上述决策规则会出现大量错误 ....

条件信息

- 使用类条件信息
  - 类条件概率$P(x \mid \omega_1)$ 和 $P(x \mid \omega_2)$ 描述了不同的类别（鲈鱼和鲑鱼）中的个体在亮度（lightness ）分布上的差异

类条件概率密度函数

令 x 是一个连续的随机变量，
$p(x \mid \omega)$ 是给定类别 ω 后 x 的概率密度.

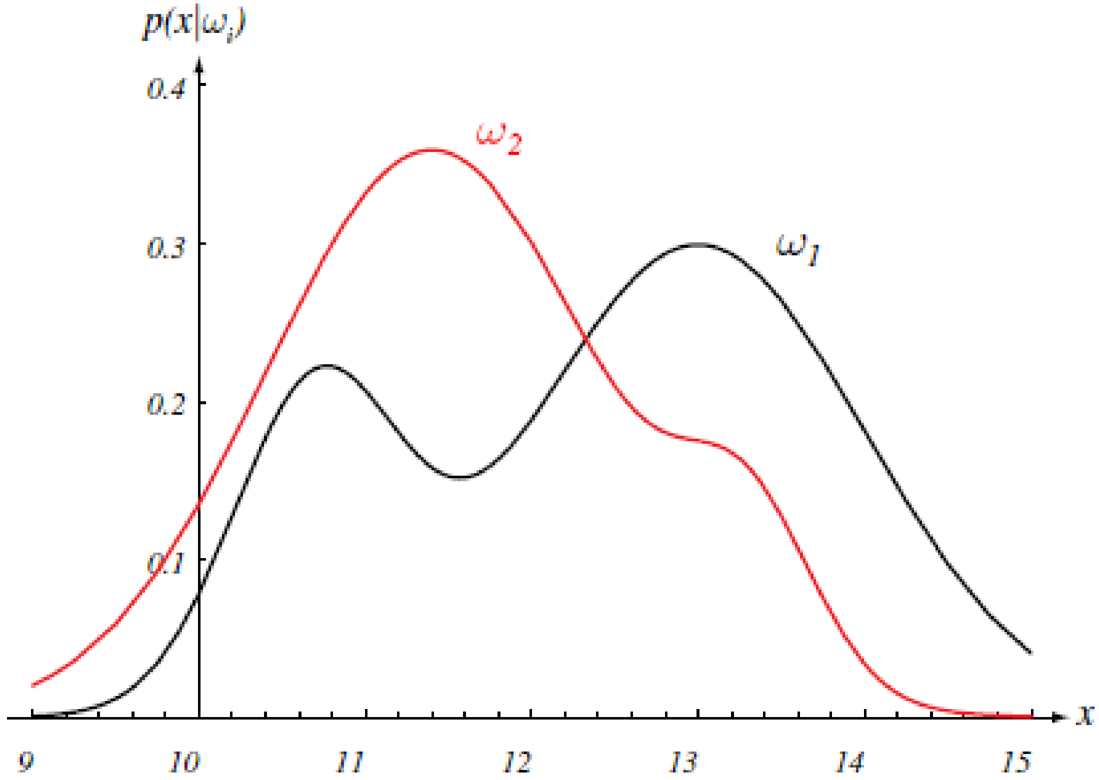
例如, 在给定类别为鲑鱼类时， 该类别中亮度不同的鱼的概率分布记为：

p(lightness | salmon) ?

或者， 在给定类别为鲈鱼类时， 该类别中亮度不同的鱼的概率分布记为：

p(lightness | sea bass) ?

条件概率密度函数



- 类条件概率密度函数（class-conditional probability density functions）表示的是在类 ω_i 中，给定的特征测量值为 x 的模式的概率密度。
- 若 x 表示鱼的亮度, 上图的两条曲线可以描述两类鱼在不同亮度下的比例。
- 在规范化的密度函数中，每条曲线下的面积都是 1.0.

后验，似然概率，证据

Posterior, likelihood, evidence

$$P(\omega_j \mid x) = P(x \mid \omega_j)\, P(\omega_j) \,/\, P(x)$$

在两类情况下

$$P(x) = \sum_{j=1}^{2} P(x \mid \omega_j) P(\omega_j)$$

$$Posterior = (Likelihood \cdot Prior) \,/\, Evidence$$

贝叶斯公式（Bayes Formula）

如何使用先验、类条件件密度获得类别状态的概率?

Bayes 公式.

likelihood

prior probability

$$P\left(\omega_j \middle| x\right) = \frac{p\left(x \middle| \omega_j\right) P\left(\omega_j\right)}{p\left(x\right)}$$

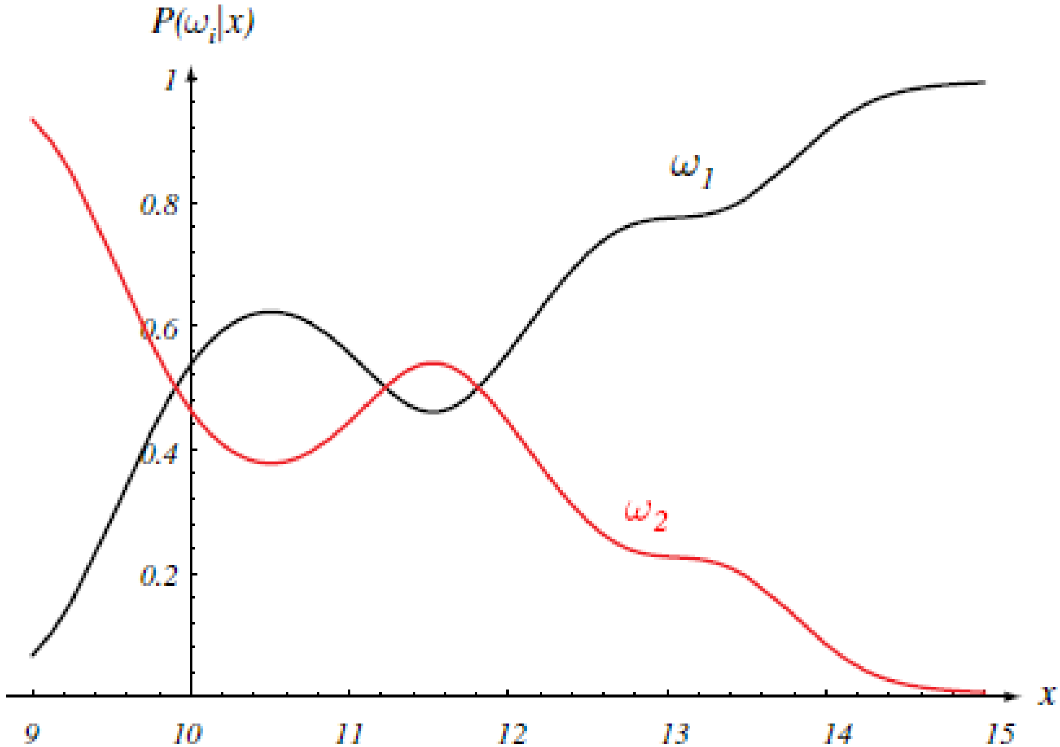
posterior probability

evidence

Bayes 决策:

若 $P\left(\omega_1 \middle| x\right) > P\left(\omega_2 \middle| x\right)$ ，则决策为 ω_1 ，否则决策为 ω_2

后验概率（Posterior probabilities）



- 如图Fig. 2.1所示是先验（priors）概率分别为$P(\omega_1) = 2/3$和$P(\omega_2) = 1/3$时，计算出的后验概率密度（Posterior probabilities ）的曲线。
- 在这种情况下，若一个模式的特征测量值为 $x= 14$, 则该模式在类ω_2中的概率约为 0.08, 在ω_1中的约为 0.92。对每个x, 后验概率之和为 1.0

Bayes 决策

决策是由后验概率给出的,

$\boldsymbol{x}$ 是一个观测数据:

若 $P(\omega_1 \mid x) > P(\omega_2 \mid x)$ ⟹ True 类状态 $= \omega_1$

若 $P(\omega_1 \mid x) < P(\omega_2 \mid x)$ ⟹ True 类状态 $= \omega_2$

这样:

当我们观测到一个特定 x 时，决策的错误概率为：

$P(error \mid x) = P(\omega_1 \mid x)$，*如果决策为* ω_2

$P(error \mid x) = P(\omega_2 \mid x)$，*如果决策为* ω_1

Bayes 规则的简化

Bayes 公式.

$$P(\omega_j|x)=\frac{p(x|\omega_j)P(\omega_j)}{p(x)}$$

证据 $p(x)$ 对所有类别状态都是相同的，在决策时可以省略。

规则:

若 $P(x|\omega_1)P(\omega_1)>P(x|\omega_2)P(\omega_2)$ 判定为类别 ω_1，否则为 ω_2

分析:

若 $P(x|\omega_1)=P(x|\omega_2)$ 则决策只依赖于先验概率.

若 $P(\omega_1)=P(\omega_2)$ 则决策只依赖于似然概率.

最小化错误概率

- 最小化错误概率

若 $P(\omega_1 \mid x) > P(\omega_2 \mid x)$ 时，决策为 ω_1 ；

否则决策为 ω_2

因此:

$$P(error \mid x) = min\,[P(\omega_1 \mid x), P(\omega_2 \mid x)]$$

(Bayes decision)

错误概率

什么是错误概率?

$$P(error|x)=\begin{cases}P(\omega_1) & if\ \ we\ decide\ \ \omega_2\\ P(\omega_2) & if\ \ we\ decide\ \ \omega_1\end{cases}$$

Bayes 规则可以使错误概率最小化吗?

$$P(error)=\int_{-\infty}^{\infty}P(error,x)dx=\int_{-\infty}^{\infty}P(error|x)p(x)dx$$

如果对每一个 x 我们都使错误最小化

$P(error|x)$ 是否可以尽可能小?

回答是 "yes".

贝叶斯决策理论 – 连续特征情况

对前面的决策思想进行一般化

① 特征的数量可使用一个以上

② 类别状态的数量可用于两类以上

③ 不仅仅做出类别状态的决策，可以允许采取某种行为

④ 引入损失函数（ loss of function ）比使用错误概率（probability of error）更具有一般性

贝叶斯决策理论 – 连续特征情况

- 允许采取某种行动，而不仅仅只是分类
  - 首先，要允许拒识的可能性

即在近似或坏的情况下，拒绝做出决策!

损失函数（Loss Function）

损失函数规定了在采取每一种行为时需要付出的代价

令

$\{\omega_1, \omega_2, \ldots, \omega_c\}$ 是具有个 c 类的类状态集合

令

$\{\alpha_1, \alpha_2, \ldots, \alpha_a\}$ 是可能采取的行为集合

令 $\lambda(\alpha_i \mid \omega_j)$ 是类状态为 ω_j 时，采取行为 α_i 而引起的损失

条件风险（Conditional Risk）

总风险

$$R = Sum\ of\ all\ \underbrace{R(\alpha_i \mid x)}_{\textbf{Conditional risk}}\ for\ i = 1,\ldots,a$$

最小化 **R** ⟺ 最小化 $R(\alpha_i \mid x)\ for\ i = 1,\ldots, a$

$$R(\alpha_i \mid x) = \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid x)$$

$for\ i = 1,\ldots,a$

Bayes 风险决策

选择行为α_i，使得风险 $R(\alpha_i \mid x)$为最小，即

$\Longrightarrow$ R 是最小的

在这种情况下的 R 称为贝叶斯风险（**Bayes risk**）

= 等价于所能得到的最好的性能!

损失函数小结

令 $\{\omega_1, \omega_2, \cdots, \omega_c\}$ 是可能的类状态.

令 $\{a_1, a_2, \cdots, a_k\}$ 是可能的行为.

损失函数

$\lambda(a_i|\omega_j)$ 是类状态为 ω_j 时采取行为 a_i 引起的损失

假设我们观测到了一个特定的 x，考虑采取行为 a_i.
如果真实的类状态是 ω_j，则损失为 $\lambda(a_i|\omega_j)$

期望损失 $R(a_i|x) = \sum_j \lambda(a_i|\omega_j) P(\omega_j|x)$

(这也称为条件风险.)

决策: 选择使条件风险最小化的行为
(**可能的最好性能 **)

两类时的风险决策问题

假设

α_1 : 在类别为 ω_1 时采取的行为

α_2 : 在类别为 ω_2 时采取的行为

$\lambda_{ij} = \lambda(\alpha_i \mid \omega_j)$ 当真实状态为时 ω_j，决策为状态 ω_i 而引起的损失

条件风险:

$$R(\alpha_1 \mid x) = \lambda_{11}P(\omega_1 \mid x) + \lambda_{12}P(\omega_2 \mid x)$$

$$R(\alpha_2 \mid x) = \lambda_{21}P(\omega_1 \mid x) + \lambda_{22}P(\omega_2 \mid x)$$

两类问题的决策

规则如下:

若 $R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$

则采取行为 α_1: 此时做出的“决策为 ω_1”

与上述规则等价的规则为:

决策为 ω_1 , 若:

$$(\lambda_{21} - \lambda_{11})\, P(x \mid \omega_1)\, P(\omega_1) > (\lambda_{12} - \lambda_{22})\, P(x \mid \omega_2)\, P(\omega_2)$$

否则，决策为 ω_2

两类情况

对该问题的另一种描述如下.

决策为 ω_1，若

$$\underbrace{\frac{p(x|\omega_1)}{p(x|\omega_2)}}_{\text{likelihood ratio}} > \underbrace{\frac{(\lambda_{12}-\lambda_{22})}{(\lambda_{21}-\lambda_{11})}\frac{p(\omega_2)}{p(\omega_1)}}_{\text{constant}}$$

描述为:

决策为 ω_1，如果似然比大于一个给定的阈值。

注意：该阈值独立于观测量 x.

似然比（Likelihood ratio）

前述规则等价于下面的规则:

$$if\ \frac{P(x|\omega_1)}{P(x|\omega_2)} > \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)}$$

则采取行为 α_1 (此时决策为 ω_1)
否则采取行为 α_2 (此时决策为 ω_2)

最优决策

若似然比大于一个独立于输入模式 x 的阈值，则决策为 ω_1是最优决策