引言

- 鲑鱼与鲈鱼的例子中：
  - 类别状态与先验
    - 类别状态为一个随机变量，本例中设ω_1表示鲈鱼，ω_2表示鲑鱼；
    - 假设获取鲑鱼或鲈鱼的先验概率相同：
      - $P(\omega_1) = P(\omega_2)$ (均匀先验)
      - $P(\omega_1) + P(\omega_2) = 1$ (排他性和穷尽性)

- 仅基于先验信息的判定准则：
  - 若$P(\omega_1) > P(\omega_2)$，则将样本判定为$\omega_1$，否则判定为$\omega_2$。
- 若进一步利用类条件概率密度：
  - $p(x|\omega_1)$ 及$p(x|\omega_2)$ 描述了两种鱼类外观上的亮度差异。

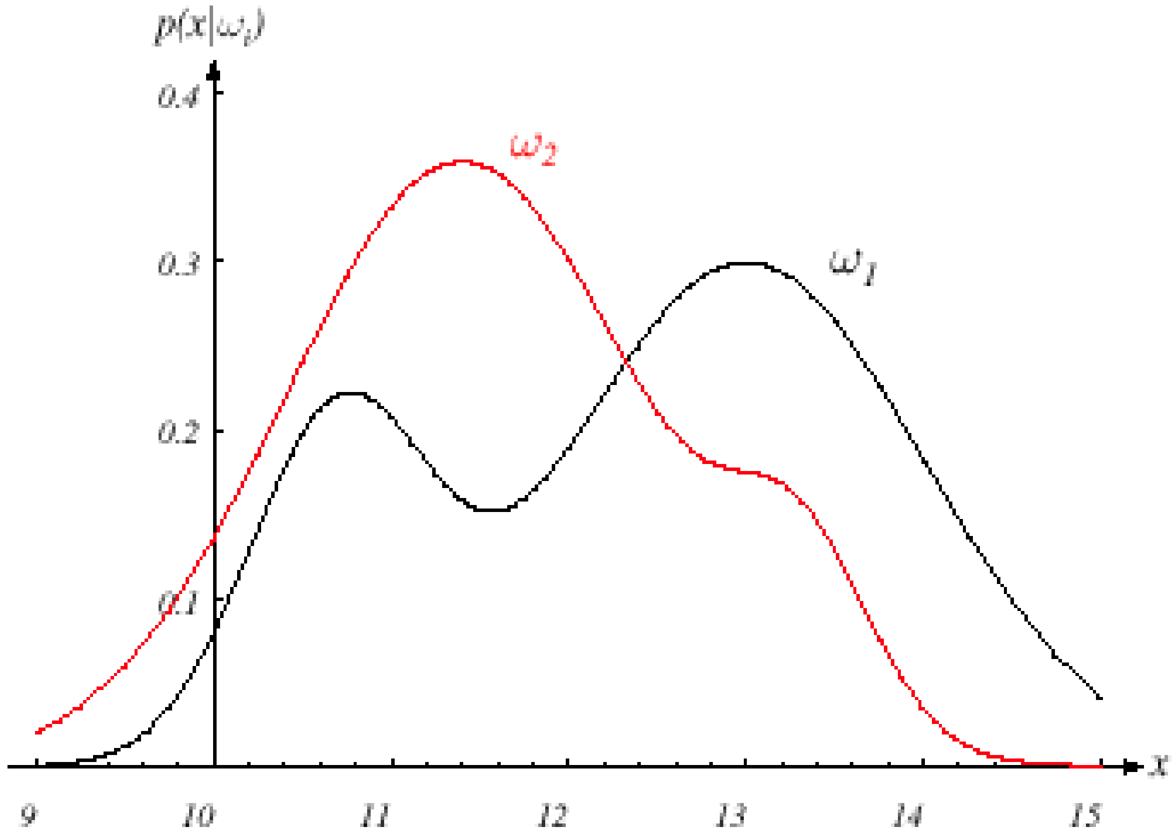

图2.1 假定的类条件概率密度函数图，显示了模式处于类别ω_i时观察某个特定特征值x的概率密度。如果x代表了鱼的亮度，那么这两条曲线可描述两种鱼的亮度区别。概率密度函数已归一化，因此每条曲线下的面积为1。

- 后验概率、似然值与证据因子
  - $P(\omega_j|x) = \left(p(x|\omega_j) \cdot P(\omega_j)\right)/p(x)$
  - 后验概率 =(似然值 · 先验概率)/证据因子
  - 其中在两类情况下：

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

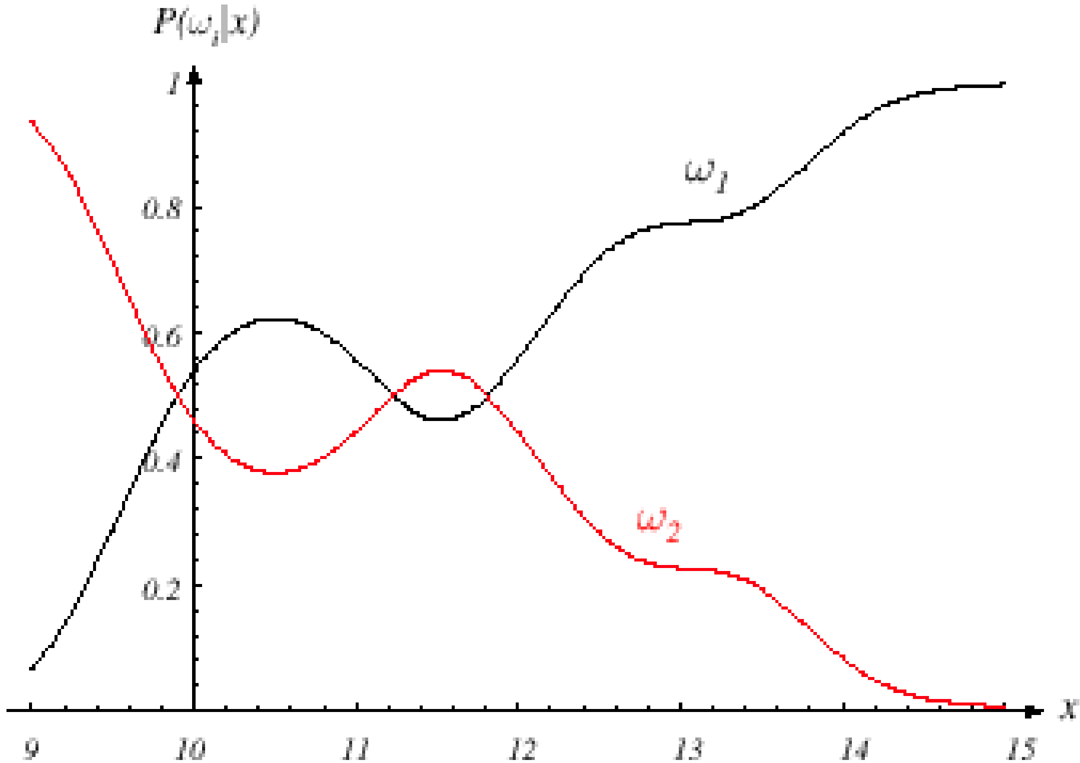

图2.2 在先验概率$P(\omega_1) = 2/3$，$P(\omega_2) = 1/3$，类条件概率密度如图2.1时对应的后验概率。此时，若给定模式的特征值为$x = 14$，则它属于ω_2类的概率约为0.08，属于ω_1的概率约为0.92。在每个x处的后验概率之和为1.0。

- 基于后验概率的决策准则

设x为观察值:

若$P(\omega_1|x) > P(\omega_2|x) \Longrightarrow$ 类别判定为ω_1

若$P(\omega_1|x) < P(\omega_2|x) \Longrightarrow$ 类别判定为ω_2

-如果将x判定为ω_2，则分类错误的概率为:

$$P(\mathrm{error}|x) = P(\omega_1|x)$$

−如果将x判定为ω_1，则分类错误的概率为:

$$P(\mathrm{error}|x) = P(\omega_2|x)$$

- 最小化错误概率:

  若$P(\omega_1|x) > P(\omega_2|x)$，则判定类别为$\omega_1$；反之，判定为$\omega_2$。

  可以证明:

$$P(\mathrm{error}|x) = \min[P(\omega_1|x), P(\omega_2|x)]$$

(贝叶斯决策准则)

贝叶斯决策论——连续特征

- 对前述思想进行推广：
  - 允许多于一个特征
  - 允许多于两种类别状态
  - 允许其它行动而不仅是判定类别
  - 引入比错误概率具有更一般意义的损失函数

- 允许类别判定以外的行动主要是允许拒绝的可能（拒绝在难以抉择的情况下做出决策！）
  - 令$\{\omega_1, \omega_2, \dots, \omega_c\}$表示$c$个类别的集合。
  - 令$\{\alpha_1, \alpha_2, \dots, \alpha_a\}$表示可能行动的集合。
- 用损失函数定义采取每项行动带来的代价：
  - 令$\lambda(\alpha_i|\omega_j)$表示当类别状态为$\omega_j$时采取行动$\alpha_i$带来的损失。

设$\mathbf{x}$为具有多个特征分量的向量，即特征向量。

- 观察到$\mathbf{x}$时采取行动α_i的条件风险：

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

- 总风险R = $\mathbf{x}$的空间中$R(\alpha_i|\mathbf{x})$的总和
  – 最小化$R \Leftrightarrow$最小化$R(\alpha_i|\mathbf{x})$，其中$i = 1,2,\ldots,a$
- 选择行动α_i，使得$R(\alpha_i|\mathbf{x})$最小化$\Longrightarrow R$被最小化，此时的总风险R称为贝叶斯风险，它是可获得的最优结果。

- 两类分类问题

  α_1：对应类别判定ω_1

  α_2：对应类别判定ω_2

  $\lambda_{ij} = \lambda(\alpha_i|\omega_j)$：表示当实际类别为$\omega_j$时，采取行动$\alpha_i$（此时即为判定类别$\omega_i$）带来的损失。

  – 贝叶斯决策规则如下：

    ○如果$R(\alpha_1|\mathbf{x}) < R(\alpha_2|\mathbf{x})$，则采取行动$\alpha_1$：判定为$\omega_1$。

此时的条件风险:

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

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

结合贝叶斯公式，等价决策规则为:

– 如果

$$(\lambda_{21} - \lambda_{11})p(\mathbf{x}|\omega_1)P(\omega_1) > (\lambda_{12} - \lambda_{22})p(\mathbf{x}|\omega_2)P(\omega_2)$$

则判定为ω_1；否则，判定为ω_2 。

前述规则可等价为下面的似然比准则：

如果

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

则采取行动α_1（判定为ω_1）；

否则，采取行动α_2（判定为ω_2）。

练习

请为$x \in \{0.8, 1.5, 0.6, 2.0\}$选择最优分类决策，其中类别空间为$\{\omega_1, \omega_2\}$。

$p(x|\omega_1)$为正态分布$\mathrm{N}(2, 0.5^2)$

$p(x|\omega_2)$为正态分布$\mathrm{N}(1.5, 0.5^2)$

$$P(\omega_1) = 2/3$$

$$P(\omega_2) = 1/3$$

$$\lambda = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix}$$

作业：2

最小错误率分类

- 此时，行动指关于类别的决策，对应的目标是寻找能最小化错误概率（即错误率）的决策规则。
  - 如果采取了行动α_i，而真实类别为ω_j，则：当$i=j$时，这个决策是正确的；当$i \neq j$时，这个决策是错误的。

- 此时，$\lambda(\alpha_i|\omega_j)$为0-1损失函数：

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \qquad i,j = 1,\ldots,c$$

条件风险简化为：

$$R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|\mathbf{x})$$
$$= \sum_{j\neq i} P(\omega_j|\mathbf{x}) = 1 - P(\omega_i|\mathbf{x})$$

该风险就是分类错误的概率。

- 为实现风险最小化（最小错误率），应最大化后验概率$P(\omega_i|\mathbf{x})$，即：
  - 当$P(\omega_i|\mathbf{x}) \geq P(\omega_j|\mathbf{x}), \forall j \neq i$时，决策为$\omega_i$。

- 更为一般损失函数下的决策区域：

  令 $\frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)} = \theta_\lambda$，则当 $\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \theta_\lambda$时，决策为$\omega_1$。

  - 如果λ是0-1损失函数，亦即$\lambda = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$，则 $\theta_\lambda = \frac{P(\omega_2)}{P(\omega_1)} = \theta_a$
  - 如果$\lambda = \begin{pmatrix} 0 & 1.2 \\ 1 & 0 \end{pmatrix}$，则$\theta_\lambda = \frac{1.2P(\omega_2)}{P(\omega_1)} = 1.2\theta_a = \theta_b$

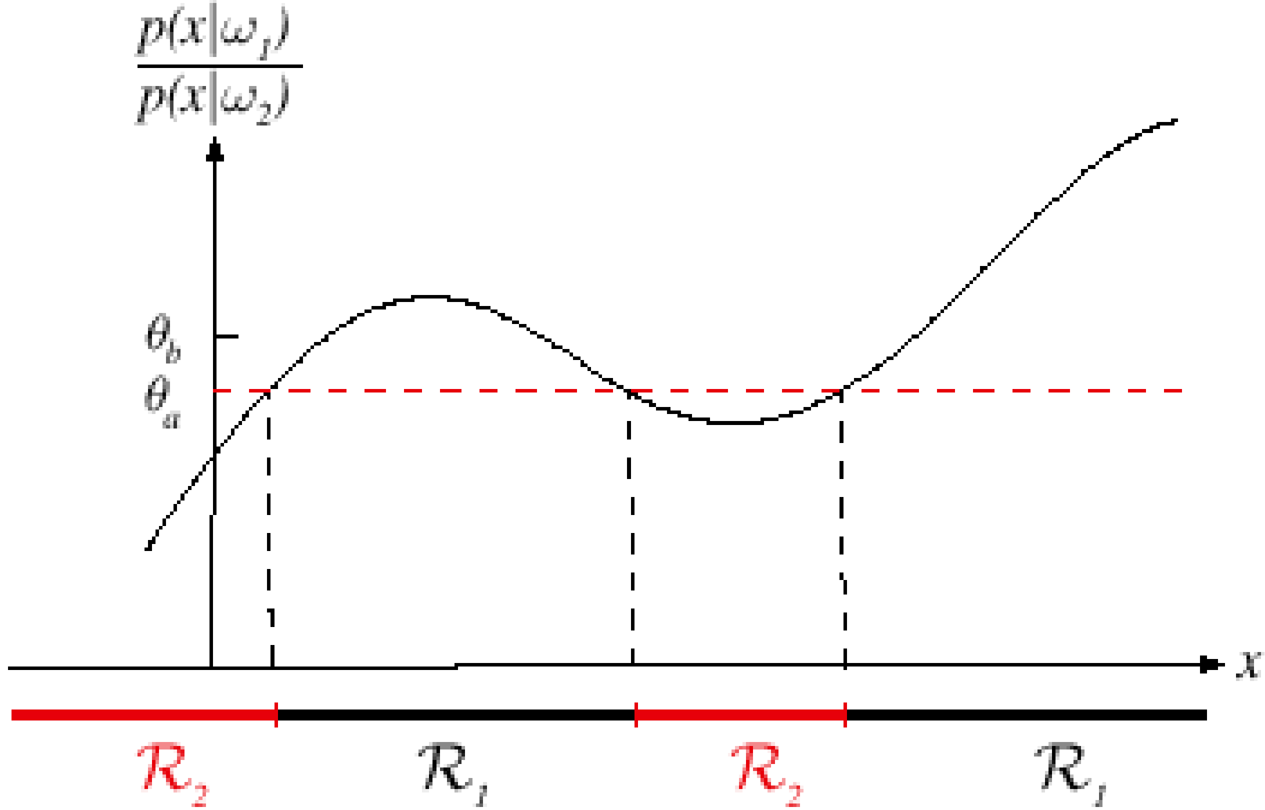

图2.3 似然比$p(x|\omega_1)/p(x|\omega_2)$（概率密度如图2.1所示）。如果我们采用0-1或分类损失，那么决策边界由阈值θ_a决定。如果该损失函数对于将ω_2错分为ω_1的惩罚比将ω_1错分为ω_2的惩罚更重，那么阈值会增大至θ_b，从而$\mathcal{R}_1$变小。