模式识别课程

贝叶斯分类器

张俊超

中南大学
航空航天学院

模式识别-贝叶斯分类器

- 模式识别算法
  - 统计模式识别
    - 线性分类器
      - 感知器
      - LMSE算法
      - SVM
      - Fisher线性判别
    - 贝叶斯分类器
      - 最小错误率
      - 最小风险
    - 最近邻分类器
      - 模板匹配
      - K近邻算法
    - 神经网络分类器
      - 多层感知器
      - CNN
    - 统计聚类算法
      - 层次聚类
      - 动态聚类
    - 特征选择与特征提取
  - 结构模式识别

模式识别-贝叶斯分类器

本章主要内容：

主要讨论**有监督的分类问题**

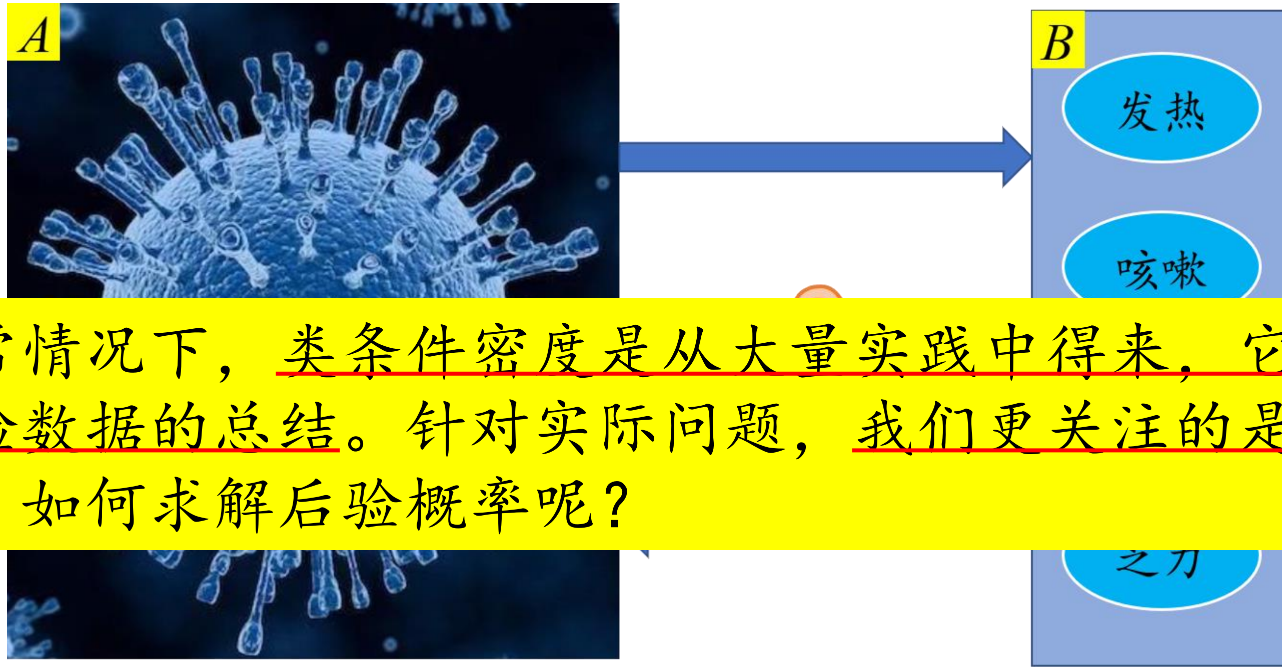
◆最小错误率贝叶斯决策

◆最小风险贝叶斯决策

◆正态分布概率模型下的最小错误率贝叶斯决策

◆概率密度函数的估计

模式识别-贝叶斯分类器



2019-nCoV　　　　症状

通常情况下，类条件密度是从大量实践中得来，它是一种经验数据的总结。针对实际问题，我们更关注的是后验概率。如何求解后验概率呢？

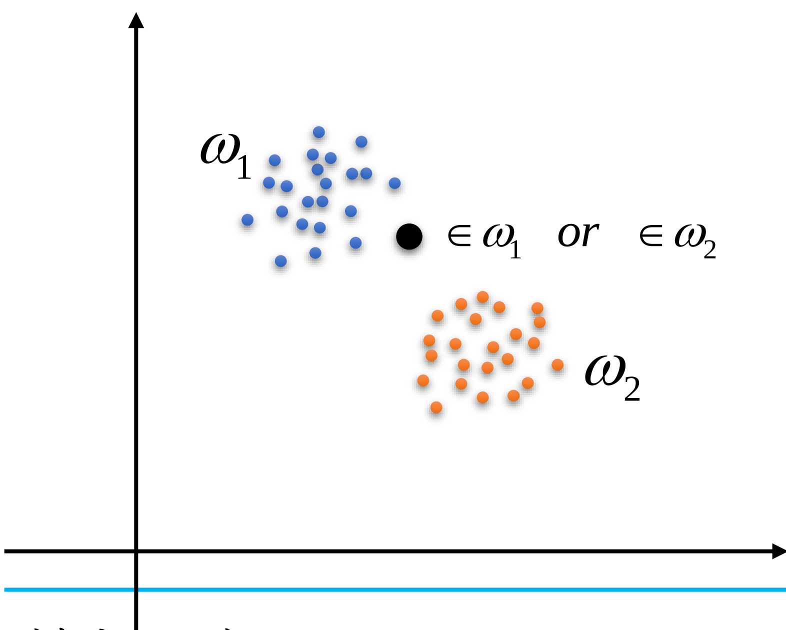
- 患病(结果)→症状(条件)的概率：$P(B|A)$　类条件密度
- 症状(条件)→患病(结果)的概率：$P(A|B)$　后验概率

模式识别-贝叶斯分类器

贝叶斯公式：

后验概率　　类条件密度　　先验概率

$$P(A|B)=\frac{P(A,B)}{P(B)}=\frac{P(B|A)P(A)}{P(B)}$$



$$P(\omega_1|\mathbf{x}) \quad vs \quad P(\omega_2|\mathbf{x})$$

模式识别-贝叶斯分类器

如果在路上我们看到一个人的背影，发现ta有长头发，那么ta是男士还是女士呢？

Who?

P(男士|长发) 和 P(女士|长发)

模式识别-贝叶斯分类器

求P(男士|长发)和P(女士|长发):

假设:

- P(长发|男士)=0.05, P(长发|女士)=0.7
- P(男士)=P(女士)=0.5

则: P(女士|长发) >P(男士|长发), ta是女士

$$P(\text{男士}\mid\text{长发})=\frac{P(\text{长发}\mid\text{男士})*P(\text{男士})}{P(\text{长发}\mid\text{男士})*P(\text{男士})+P(\text{长发}\mid\text{女士})*P(\text{女士})}=0.067$$

$$P(\text{女士}\mid\text{长发})=\frac{P(\text{长发}\mid\text{女士})*P(\text{女士})}{P(\text{长发}\mid\text{男士})*P(\text{男士})+P(\text{长发}\mid\text{女士})*P(\text{女士})}=0.933$$

模式识别-贝叶斯分类器

先验概率变化了

假设：

- P(长发|男士)=0.05， P(长发|女士)=0.7
- P(男士)=0.8，P(女士)=0.2

则：

$$P(男士|长发)=\frac{P(长发|男士)*P(男士)}{P(长发|男士)*P(男士)+P(长发|女士)*P(女士)}=0.22$$

$$P(女士|长发)=\frac{P(长发|女士)*P(女士)}{P(长发|男士)*P(男士)+P(长发|女士)*P(女士)}=0.78$$

贝叶斯分类器–最小错误率决策

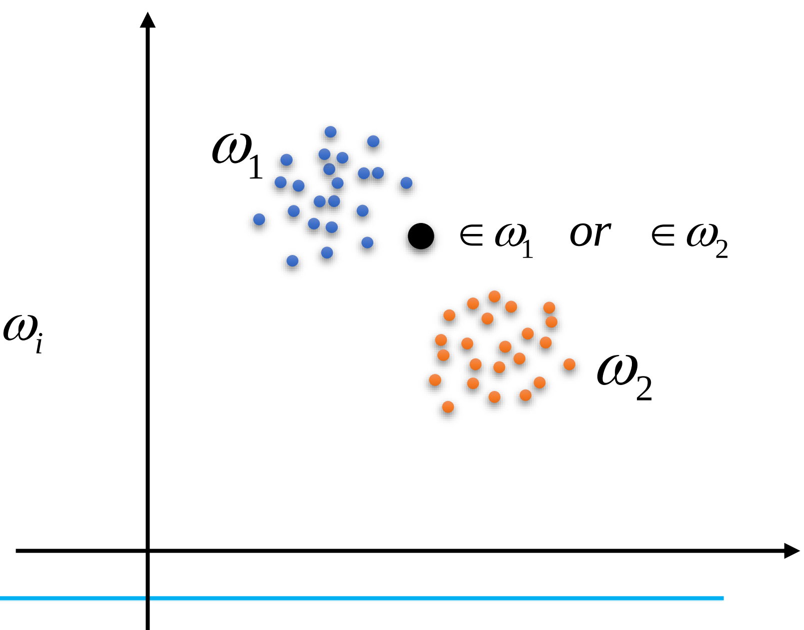
- 最大后验概率(最小错误率)的贝叶斯分类器

决策规则： 若$P(\omega_i \mid \mathbf{x}) = \max\limits_{j=1,2,\ldots,c} P(\omega_j \mid \mathbf{x})$，则$\mathbf{x} \in \omega_i$

最大后验概率决策会存在误分类吗？

$$P(\omega_i \mid \mathbf{x}) = \frac{P(\mathbf{x} \mid \omega_i) P(\omega_i)}{P(\mathbf{x})}$$

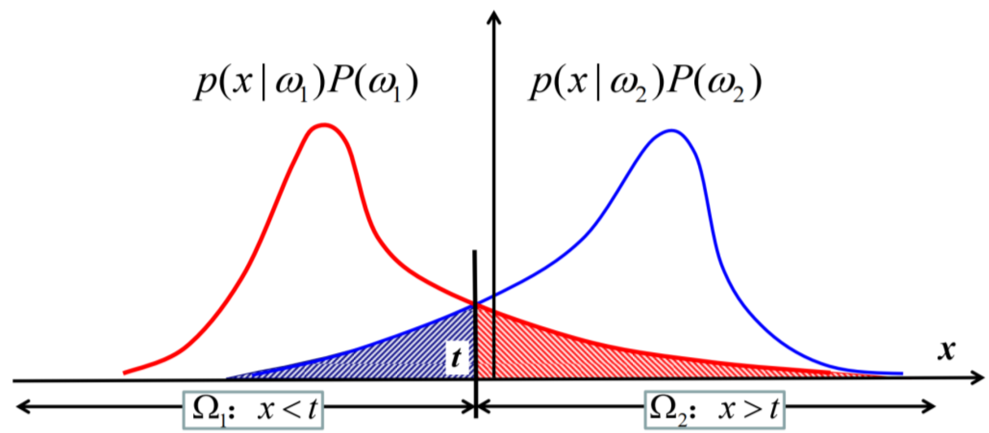
若$P(\omega_i \mid \mathbf{x}) = \max\limits_{j=1,2,\ldots,c} P(\mathbf{x} \mid \omega_j) P(\omega_j)$，则$\mathbf{x} \in \omega_i$



贝叶斯分类器－最小错误率决策

- 误分类
- 错误率：
- 正确率：

$$P(c)=1-P(e)$$



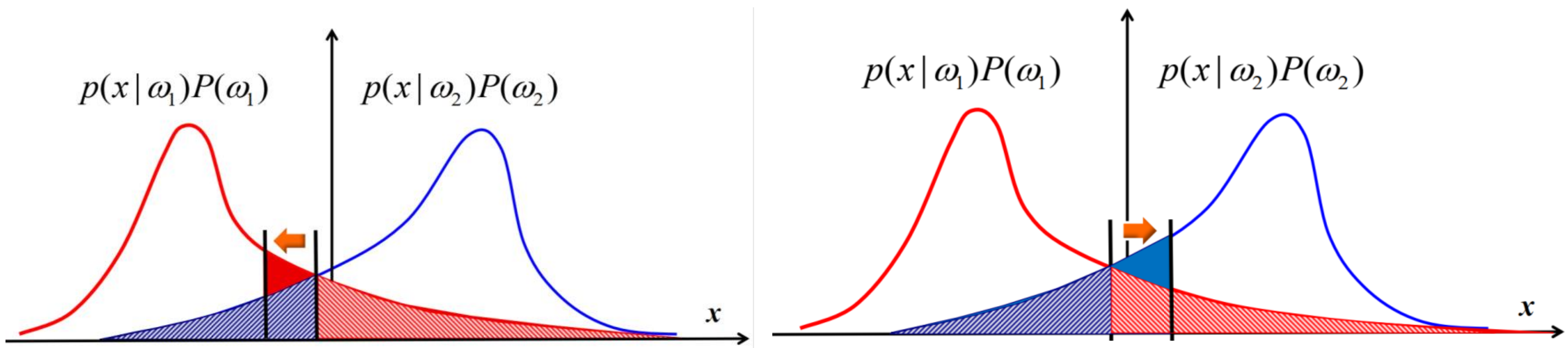
$$\begin{aligned}
P(e) &= P(x在\Omega_2中|\omega_1)+P(x在\Omega_1中|\omega_2) \\
&= \int_{\Omega_2} \mathrm{P}(\omega_1)\mathrm{p}(\mathrm{x}|\omega_1)\mathrm{dx} + \int_{\Omega_1} \mathrm{P}(\omega_2)\mathrm{p}(\mathrm{x}|\omega_2)\mathrm{dx} \\
&= \mathrm{P}(\omega_1)\boxed{\int_{\Omega_2} \mathrm{p}(\mathrm{x}|\omega_1)\mathrm{dx}} + \mathrm{P}(\omega_2)\boxed{\int_{\Omega_1} \mathrm{p}(\mathrm{x}|\omega_2)\mathrm{dx}} \\
&= \mathrm{P}(\omega_1)\mathrm{P}_1(\mathrm{e}) + \mathrm{P}(\omega_2)\mathrm{P}_2(\mathrm{e})
\end{aligned}$$

ω_1类样本决策为 ω_2类的错误率

贝叶斯分类器-最小错误率决策

为什么最大后验概率决策就是最小错误率决策呢？

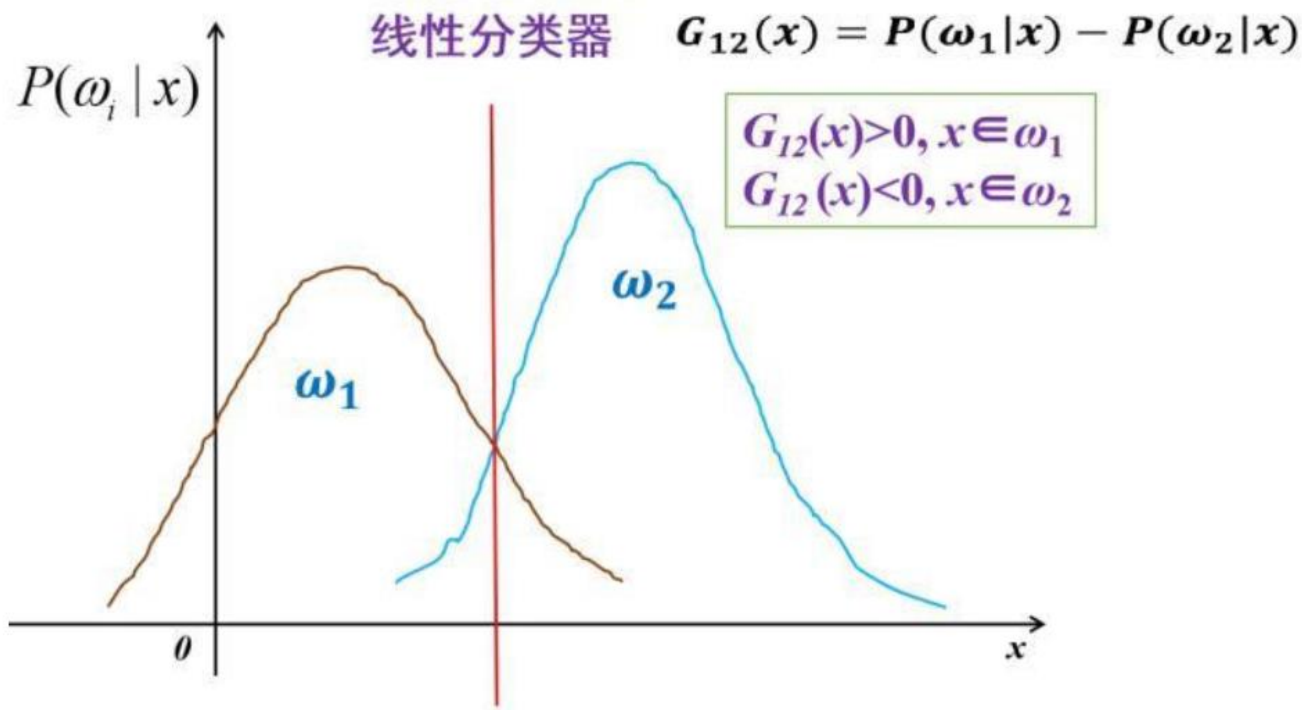
$$P(e)=\int_{\Omega_2}P(\omega_1)p(x\,|\,\omega_1)dx+\int_{\Omega_1}P(\omega_2)p(x\,|\,\omega_2)dx$$



贝叶斯分类器-最小错误率决策

最小错误率贝叶斯分类器的分类边界是什么样的呢？

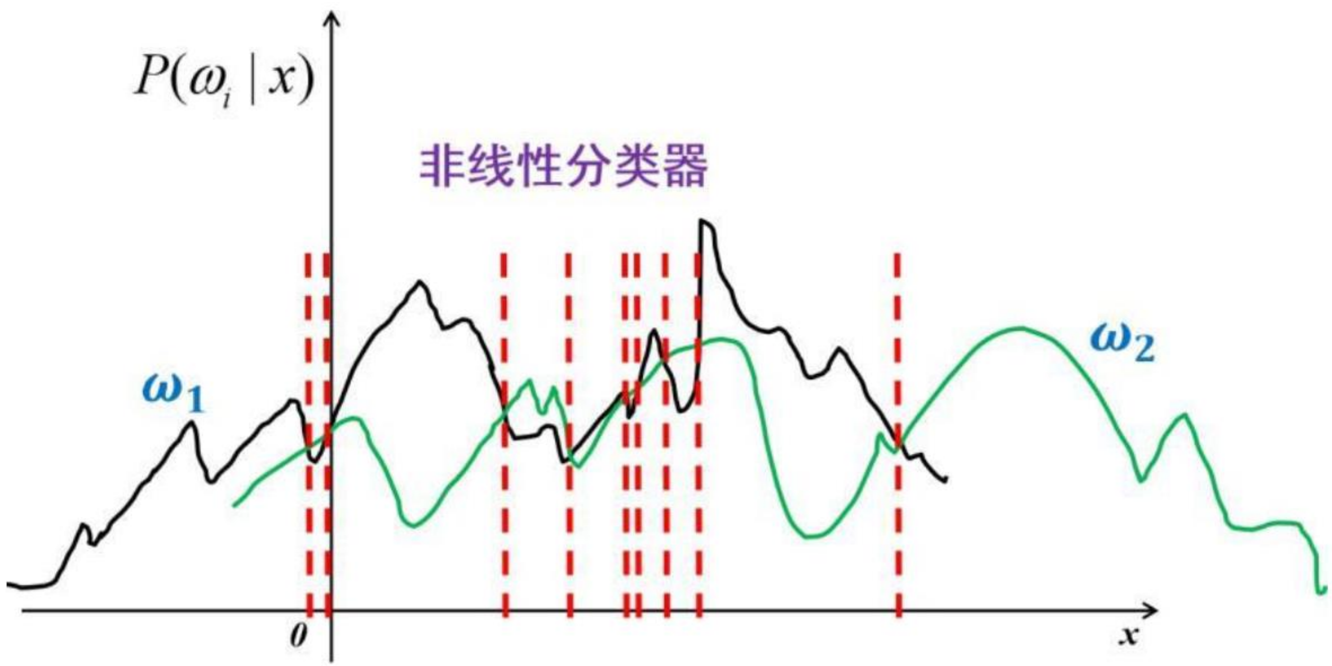
决策边界为： $P\left(\omega_i \mid \mathbf{x}\right)=P\left(\omega_j \mid \mathbf{x}\right)$



贝叶斯分类器-最小错误率决策

最小错误率贝叶斯分类器的分类边界一定是线性的吗？

分类决策边界不一定是线性的，也不一定是连续的。

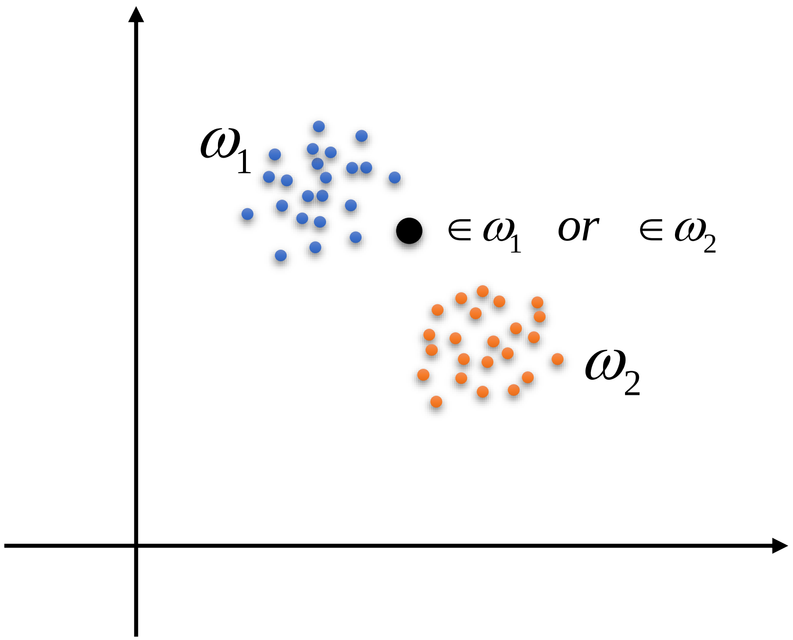


贝叶斯分类器–最小错误率决策

如何训练最小错误率贝叶斯分类器，使其实现分类呢？

$$若P(\omega_i \mid \mathbf{x}) = \max_{j=1,2,\ldots,c} P(\mathbf{x}|\omega_j)P(\omega_j)，则\mathbf{x} \in \omega_i$$

- 根据已知样本估计类条件密度和先验概率(书中第3章的内容)
- 计算待测样本的后验概率
- 根据最大后验概率决策，进行分类



贝叶斯分类器－最小错误率决策

举例：[数据真实性不具参考意义，纯为说明问题]

2019年底，新型冠状病毒(2019-nCoV)的发病率为两万分之一，普通流感的发病率可高达30%。易感人群中99%的人感染2019-nCoV病例曾出现过发热、咳嗽等急性呼吸道感染症状，而同样的易感人群中80%的普通流感患者也出现过同样症状。

Q: 现有一位患者属于易感人群，并出现了发热、咳嗽等急性呼吸道感染症状，请问是否应当按照人2019-nCoV疑似病例对待？

贝叶斯分类器-最小错误率决策

$$P(2019-nCoV)=\frac{0.00005}{0.00005+0.3}=0.00017$$

$$P(\text{普通流感})=0.99983$$

$$P(\text{症状}|2019-nCoV)=0.99$$

$$P(\text{症状}|\text{普通流感})=0.8$$

这样做，合理吗？如果把2019-nCoV误诊为普通流感，后果会怎样？

$$P(2019-nCoV|\text{症状})=\frac{0.99*0.00017}{0.99*0.00017+0.8*0.99983}=0.0002$$

$$P(\text{普通流感}|\text{症状})=0.9998$$

按照最大后验概率决策的话，该患者不应当按照2019-nCoV疑似病例对待。

贝叶斯分类器–最小风险贝叶斯决策

- 如果把普通流感误诊为2019-nCoV，后果会怎样？
- 如果把2019-nCoV误诊为普通流感，后果会怎样？

社会风险哪个更大？

- 仅仅考虑最小错误率是不够的
- 还应当把所采取的分类决策所带来的后果考虑进去

最小风险贝叶斯分类器

贝叶斯分类器-最小风险贝叶斯决策

概念：

- 决策：$\alpha_1,\alpha_2,\dots,\alpha_k$

  α_i：把待识别样本归到 ω_i

- 损失函数：实际状态为 ω_j，采取 α_i 决策

$$\lambda(\alpha_i,\omega_j), i=1,2,\dots,k,\quad j=1,2,\dots,c$$

贝叶斯分类器－最小风险贝叶斯决策

损失函数通常以表格的形式给出(决策表)

表 2-2-1

损失 \ 状态 决策	ω_1	ω_2	$\cdots$	ω_j	$\cdots$	ω_m
α_1	$\lambda(\alpha_1,\omega_1)$	$\lambda(\alpha_1,\omega_2)$	$\cdots$	$\lambda(\alpha_1,\omega_j)$	$\cdots$	$\lambda(\alpha_1,\omega_m)$
α_2	$\lambda(\alpha_2,\omega_1)$	$\lambda(\alpha_2,\omega_2)$	$\cdots$	$\lambda(\alpha_2,\omega_j)$	$\cdots$	$\lambda(\alpha_2,\omega_m)$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
α_i	$\lambda(\alpha_i,\omega_1)$	$\lambda(\alpha_i,\omega_2)$	$\cdots$	$\lambda(\alpha_i,\omega_j)$	$\cdots$	$\lambda(\alpha_i,\omega_m)$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
α_a	$\lambda(\alpha_a,\omega_1)$	$\lambda(\alpha_a,\omega_2)$	$\cdots$	$\lambda(\alpha_a,\omega_j)$	$\cdots$	$\lambda(\alpha_a,\omega_m)$

贝叶斯分类器-最小风险贝叶斯决策

- 采取决策α_i的期望损失为：

$$R(\alpha_i \mid \mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i, \omega_j) P(\omega_j \mid \mathbf{x}), i = 1, 2, \ldots, k$$

- 最小风险贝叶斯决策：

$$\text{若} R(\alpha_i \mid \mathbf{x}) = \min_{j=1,2,\ldots,k} R(\alpha_j \mid \mathbf{x}), \text{ 则} \alpha = \alpha_i$$

- 针对二分类的情况：

$$\begin{cases} \lambda_{11} P(\omega_1 \mid \mathbf{x}) + \lambda_{12} P(\omega_2 \mid \mathbf{x}) < \lambda_{21} P(\omega_1 \mid \mathbf{x}) + \lambda_{22} P(\omega_2 \mid \mathbf{x}), \mathbf{x} \in \omega_1 \\ \lambda_{11} P(\omega_1 \mid \mathbf{x}) + \lambda_{12} P(\omega_2 \mid \mathbf{x}) > \lambda_{21} P(\omega_1 \mid \mathbf{x}) + \lambda_{22} P(\omega_2 \mid \mathbf{x}), \mathbf{x} \in \omega_2 \end{cases}$$

其中，$\lambda_{ij} = \lambda(\alpha_i, \omega_j)$

贝叶斯分类器–最小风险贝叶斯决策

二分类问题：当$\lambda_{11}=\lambda_{22}=0,\lambda_{12}=\lambda_{21}=1$时

$$\begin{cases}\lambda_{11}P(\omega_1|\mathbf{x})+\lambda_{12}P(\omega_2|\mathbf{x})<\lambda_{21}P(\omega_1|\mathbf{x})+\lambda_{22}P(\omega_2|\mathbf{x}),\mathbf{x}\in\omega_1\\ \lambda_{11}P(\omega_1|\mathbf{x})+\lambda_{12}P(\omega_2|\mathbf{x})>\lambda_{21}P(\omega_1|\mathbf{x})+\lambda_{22}P(\omega_2|\mathbf{x}),\mathbf{x}\in\omega_2\end{cases}$$

其中，$\lambda_{ij}=\lambda(\alpha_i,\omega_j)$

0-1决策表

$$\begin{cases}P(\omega_2|\mathbf{x})<P(\omega_1|\mathbf{x}),\mathbf{x}\in\omega_1\\ P(\omega_2|\mathbf{x})>P(\omega_1|\mathbf{x}),\mathbf{x}\in\omega_2\end{cases}$$

最小风险贝叶斯决策⟶最大后验概率决策

贝叶斯分类器–最小风险贝叶斯决策

决策	状态	
	2019-nCoV	普通流感
2019-nCoV	0	1
普通流感	10000	0

$P\left(2019-nCoV|\text{症状}\right)=0.0002$ $P\left(\text{普通流感}|\text{症状}\right)=0.9998$

$$R\left(2019-nCoV|\text{症状}\right)=0*P\left(2019-nCoV|\text{症状}\right)+1*P\left(\text{普通流感}|\text{症状}\right)$$
$$=0.9998$$
$$R\left(\text{普通流感}|\text{症状}\right)=10000*P\left(2019-nCoV|\text{症状}\right)+0*P\left(\text{普通流感}|\text{症状}\right)$$
$$=2$$

按照最小风险贝叶斯决策的话，该患者应当按照2019-nCoV疑似病例对待。

模式识别-贝叶斯分类器

- 小结
  - 贝叶斯分类器
    1. 最小错误率决策(错误率最小化)
    2. 最小风险决策(风险最小化)
  - 最小风险决策需要事先指定决策表，决策表不同分类结果也不同。
  - 需要事先知道(估计)类条件密度和先验概率。

先验概率：可以从大量重复实验或数据统计中得到

类条件密度估计：估计各个特征维度的联合概率分布。若不独立，很难估计。一般假设各个特征维度相互独立，进而估计联合概率分布，这就是“朴素贝叶斯分类器”。

统计决策方法-ROC曲线

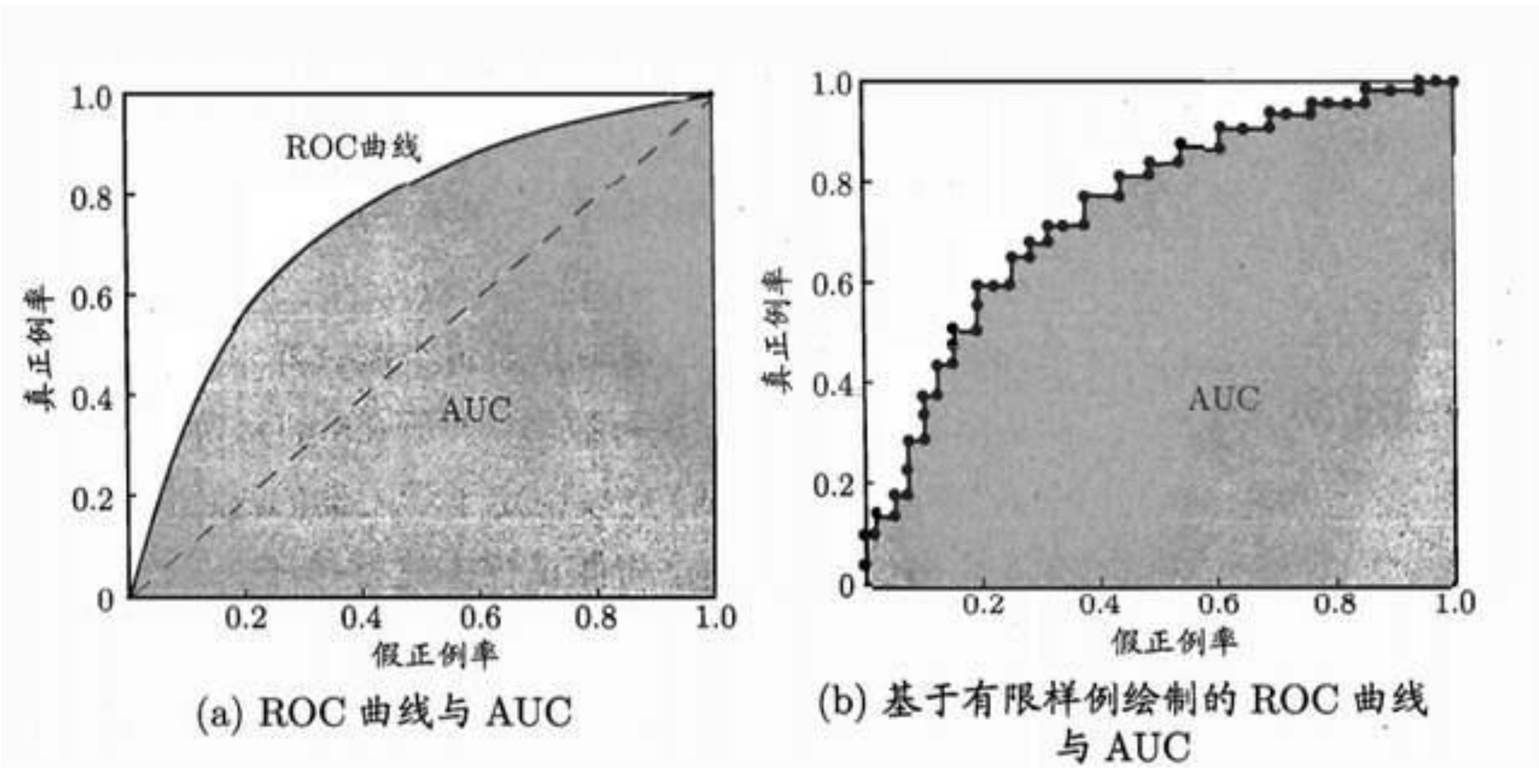
- **真阳性**(True Positive Rate)：阳性样本被预测正确的比例
- **假阳性**(False Positive Rate)：阴性样本被预测错误的比例

真实值	预测值	
	阳性	阴性
阳性	TP	FN
阴性	FP	TN

$$TPR = \frac{TP}{TP + FN} \qquad FPR = \frac{FP}{FP + TN}$$

统计决策方法-ROC曲线

AUC(area under ROC curves)：越接近1，性能越好



统计决策方法-ROC曲线

举例：

样本类别：P(positive)和N(Negative)

样本	预测属于P的概率	真实类别
1	0.1	N
2	0.35	P
3	0.4	N
4	0.8	P

统计决策方法-ROC曲线

- 截断点为0.1(概率>=0.1，就判别为P类)

样本	预测属于P的概率	真实类别
1	0.1	N
2	0.35	P
3	0.4	N
4	0.8	P

真实值	预测值	
	阳性	阴性
阳性	TP=2	FN=0
阴性	FP=2	TN=0

$$TPR = \frac{TP}{TP + FN} = 1$$

$$FPR = \frac{FP}{FP + TN} = 1$$

统计决策方法-ROC曲线

- 截断点为0.35(概率>=0.35，就判别为P类)

样本	预测属于P的概率	真实类别
1	0.1	N
2	0.35	P
3	0.4	N
4	0.8	P

真实值	预测值	
	阳性	阴性
阳性	TP=2	FN=0
阴性	FP=1	TN=1

$$TPR = \frac{TP}{TP + FN} = 1$$

$$FPR = \frac{FP}{FP + TN} = 0.5$$

统计决策方法-ROC曲线

- 截断点为0.4(概率>=0.4，就判别为P类)

样本	预测属于P的概率	真实类别
1	0.1	N
2	0.35	P
3	0.4	N
4	0.8	P

真实值	预测值	
	阳性	阴性
阳性	TP=1	FN=1
阴性	FP=1	TN=1

$$TPR = \frac{TP}{TP + FN} = 0.5$$

$$FPR = \frac{FP}{FP + TN} = 0.5$$

统计决策方法-ROC曲线

- 截断点为0.8(概率>=0.8，就判别为P类)

样本	预测属于P的概率	真实类别
1	0.1	N
2	0.35	P
3	0.4	N
4	0.8	P

真实值	预测值	
	阳性	阴性
阳性	TP=1	FN=1
阴性	FP=0	TN=2

$$TPR = \frac{TP}{TP + FN} = 0.5$$

$$FPR = \frac{FP}{FP + TN} = 0$$

统计决策方法-ROC曲线

- 截断点为0.9(概率>=0.9，就判别为P类)

样本	预测属于P的概率	真实类别
1	0.1	N
2	0.35	P
3	0.4	N
4	0.8	P

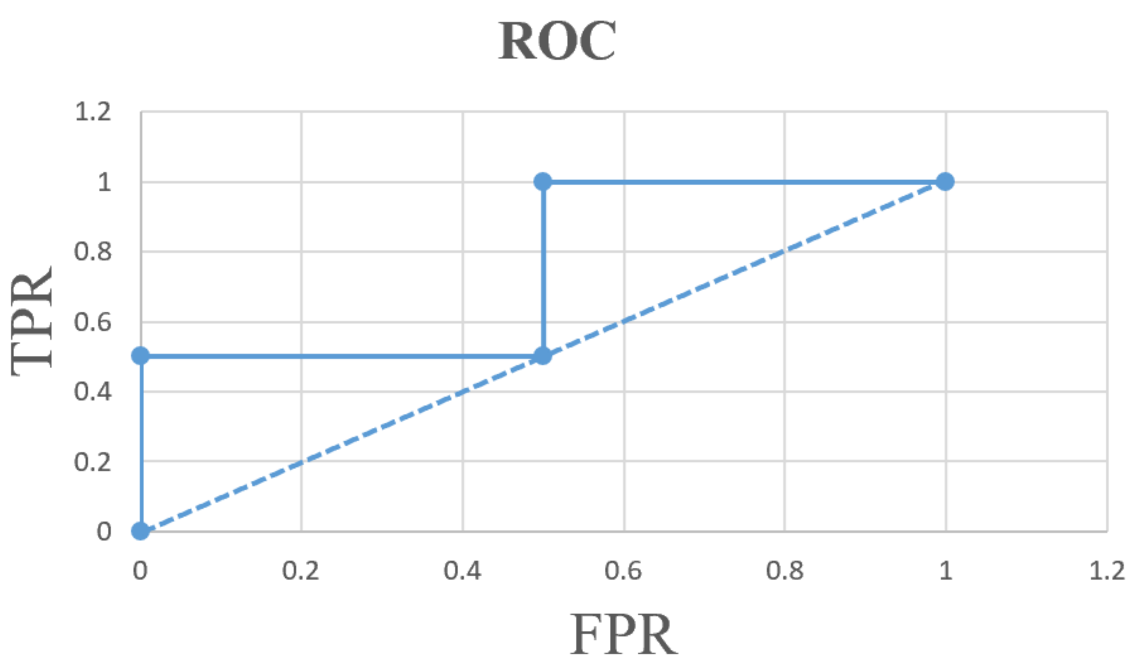
真实值	预测值	
	阳性	阴性
阳性	TP=0	FN=2
阴性	FP=0	TN=2

$$TPR = \frac{TP}{TP + FN} = 0$$

$$FPR = \frac{FP}{FP + TN} = 0$$

统计决策方法-ROC曲线

TPR	0	0.5	0.5	1	1
FPR	0	0	0.5	0.5	1



统计决策方法-作业

1 已知两个一维模式类别的类概率密度函数为

$$p(x|\omega_1)=\begin{cases} x, & 0\le x<1 \\ 2-x, & 1\le x\le 2 \\ 0, & else \end{cases}$$

$$p(x|\omega_2)=\begin{cases} x-1, & 1\le x<2 \\ 3-x, & 2\le x\le 3 \\ 0, & else \end{cases}$$

先验概率分别为 $p(\omega_1)=0.4, p(\omega_2)=0.6$。试求最大后验概率判决函数以及总的分类错误概率 $P(e)$。

统计决策方法-作业

2 在图像识别中，假定有灌木丛和坦克两种类型，它们的先验概率分别是 0.8 和 0.2，损失函数如下表所示，其中 ω_1 和 ω_2 分别表示灌木丛和坦克，α_1 和 α_2 表示判决为灌木丛和坦克，α_3 表示拒绝判决。

	ω_1	ω_2
α_1	0.5	6
α_2	2	1
α_3	1.5	1.5

现在做了三次实验，从类概率密度函数曲线上查得三个样本 $\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3$ 的类概率密度值如下：

$$\boldsymbol{X}_1: p(\boldsymbol{X}_1\,|\,\omega_1)=0.1,\ p(\boldsymbol{X}_1\,|\,\omega_2)=0.7$$
$$\boldsymbol{X}_2: p(\boldsymbol{X}_2\,|\,\omega_1)=0.3,\ p(\boldsymbol{X}_2\,|\,\omega_2)=0.45$$
$$\boldsymbol{X}_3: p(\boldsymbol{X}_3\,|\,\omega_1)=0.6,\ p(\boldsymbol{X}_3\,|\,\omega_2)=0.5$$

统计决策方法-作业

（1）试用贝叶斯最小误判概率准则判决三个样本各属于哪一个类型。

（2）假定只考虑前两种判决，试用贝叶斯最小风险判决准则判决三个样本各属于哪一个类

（3）把拒绝判决考虑在内，重新考核三次实验的结果。


若干素材取自网络，特此致谢！


谢谢聆听！