贝叶斯决策理论方法是统计模型决策中的一个基本方法

§2.1 最小错误率贝叶斯决策

统计概率

先验概率：据历史资料或主观判断所确定的各事件发生的概率，未经证实

条件概率：$P(A|B)=\frac{P(AB)}{P(B)}$ 条件概率公式

贝叶斯公式：通过观察B把状态的先验概率P(A)转化为后验P(A|B)

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

贝叶斯决策：在不完全情报下，对部分未知状态用主观概率估计，然后用贝叶斯公式对发生概率进行修正，再利用期望值和修正概率做出最优决策

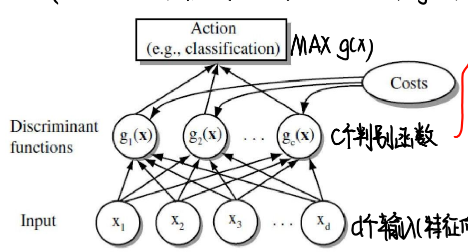
类条件概率密度函数/似然

$$P(w_i|x)=\frac{P(x|w_i)P(w_i)}{P(x)}$$ 先验概率；证据因子 $\longrightarrow \int p(x|w_i)p(w_i)dw \Rightarrow \sum_{i=0}^{N} p(x|w_i)p(w_i)$

最小错误率贝叶斯决策 若 $P(w_i|x)=\max_{j=1,\cdots,c} P(w_j|x)$，则 x 属于 w_i 类

※二分类问题下 $\begin{cases} P(w_1|x)>P(w_2|x), & x属w_1 \\ P(w_1|x)<P(w_2|x), & x属w_2 \end{cases}$ 四种等价规则 $\begin{cases} ① P(w_1|x)>P(w_2|x) & 后验概率判别 \\ ② P(x|w_1)P(w_1)>P(x|w_2)P(w_2) \\ ③ l(x)=\frac{P(x|w_1)}{P(x|w_2)}>\frac{P(w_2)}{P(w_1)} & 似然比 \to 似然比阈值 \\ ④ h(x)=\ln[l(x)]=\ln p(x|w_1)-\ln p(x|w_2)>\ln[\frac{P(w_2)}{P(w_1)}] \end{cases}$

判别函数与分类器：对多分类问题，设判别函数为 $g_i(x)$，$i=1,\cdots,c$，当 $g_i(x)>g_j(x)$，$\forall j\neq i$ 时，分类器将x归为 w_i 类



MAX g(x)

c个判别函数

d个输入(特征向量)

可设为 ① $g_i(x)=P(w_i|x)$ 后验概率 ② $g_i(x)=P(x|w_i)P(w_i)$ 类条件概率 ③ $g_i(x)=\ln P(x|w_i)P(w_i)$ 对数类条件概率

分类器计算c个判别函数，选择最大判别函数对应类别作为识别结果

决策面 方程应满足判别函数 $g_i(x)=g_j(x)$，$i\neq j$ (i与j为相邻两类) 相邻两个决策域在决策面上判别函数相等

分类的基本原理：不同模式映射点在特征空间的不同区域中散布，运用已知类别的训练样本进行学习，产生若干代数界面 $g(x)=0$，将特征空间划分为一些互不重叠的子区域

特征空间中点，直线，曲线，超曲面

※为什么是最小错误率

分类平均错误率 $P(error)=\int_{-\infty}^{\infty}P(error,x)dx=\int_{-\infty}^{\infty}P(error|x)p(x)dx$

决策规则为 $P(w_i|x)=\max_{j=1,\cdots,c}P(w_j|x)$，x属 w_i 类

$\Rightarrow P(error|x)=1-\max_{j=1,\cdots,c}P(w_j|x)$ 对任一个x可确保 P(error|x) 最小，此时 P(error) 最小

优点：对调查结果的可能性加以数量化的评价，输出非二值，而是一定置信度；巧妙地将先验知识与状态观察这两种信息有机结合起来

缺点：需要的数据多，分析计算比较复杂；先验概率起决定作用，未考虑错误分类后果

§2.2 最小风险贝叶斯

实际工作中不能仅仅考虑错误率，引入一个与损失有关联而更广泛概念——风险

条件风险 $R(\alpha_i|x)=E[\lambda(\alpha_i,w_j)]$

将 $x\in w_j$ 样本归为 w_i 的损失，i=j时判断正确

将x归为 w_i 类的决策

有决策表

对给定x采用决策 α_i，损失λ从c个 $\lambda(\alpha_i,w_j)$ 中任意选择，其相应概率 $P(w_j|x)$，有：

	ω_1	ω_2	…	ω_j	…	ω_c
α_1	$\lambda(\alpha_1,\omega_1)$	$\lambda(\alpha_1,\omega_2)$	…	$\lambda(\alpha_1,\omega_j)$	…	$\lambda(\alpha_1,\omega_c)$
α_2	$\lambda(\alpha_2,\omega_1)$	$\lambda(\alpha_2,\omega_2)$	…	$\lambda(\alpha_2,\omega_j)$	…	$\lambda(\alpha_2,\omega_c)$
…	…	…	…	…	…	…
α_i	$\lambda(\alpha_i,\omega_1)$	$\lambda(\alpha_i,\omega_2)$	…	$\lambda(\alpha_i,\omega_j)$	…	$\lambda(\alpha_i,\omega_c)$
…	…	…	…	…	…	…
α_a	$\lambda(\alpha_a,\omega_1)$	$\lambda(\alpha_a,\omega_2)$	…	$\lambda(\alpha_a,\omega_j)$	…	$\lambda(\alpha_a,\omega_c)$

$$R(\alpha_i|x)=E[\lambda(\alpha_i,w_j)]=\sum_{j=1}^{c}\lambda(\alpha_i,w_j)P(w_j|x) \quad (i=1,2,\cdots,c)$$

$P(w_j),P(x|w_j)$ $\xrightarrow{Bayes}$ $P(w_j|x)$ $\xrightarrow{决策表}$ $R(\alpha_j|x)$ $\xrightarrow{最小}$ 决策 α_k

若 $R(\alpha_k|x)=\min_{i=1,2,\cdots,a}R(\alpha_i|x)$，则对应决策 $\alpha=\alpha_k$ 判定 $x\in w_k$ 类

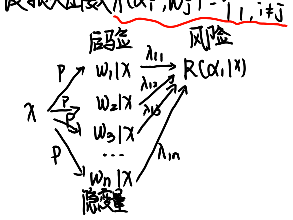
x被判为 w_k 类时的平均损失，不同决策其条件风险 $R(\alpha_i|x)$ 大小不同

※二分类问题下 $\begin{cases} R(\alpha_1|x)=\lambda(\alpha_1,w_1)P(w_1|x)+\lambda(\alpha_1,w_2)P(w_2|x) \\ R(\alpha_2|x)=\lambda(\alpha_2,w_1)P(w_1|x)+\lambda(\alpha_2,w_2)P(w_2|x) \end{cases}$，三种等价规则 $\begin{cases} ① R(\alpha_1|x)<R(\alpha_2|x) \\ ② (\lambda_{21}-\lambda_{11})P(w_1|x)>(\lambda_{12}-\lambda_{22})P(w_2|x) \\ ③ l(x)=\frac{P(x|w_1)}{P(x|w_2)}>\frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}\frac{P(w_2)}{P(w_1)} \end{cases}$

☆两种决策方法之间的关系 ⇒基于最小错误率决策是基于最小风险决策的特例

设损失函数 $\lambda(\alpha_i,w_j)=\begin{cases}0, & i=j\\ 1, & i\neq j\end{cases}$ 最小错误率决策退化为最小风险决策

0-1损失函数



§2.3 Neyman-Pearson 决策与两类错误率

1、两类错误率

状态（真实）	决策（预测）	
	阳性	阴性
阳性	真阳性（TP）	假阴性（FN）
阴性	假阳性（FP）	真阴性（TN）

假阳性率 假阳性样本占总阴性样本比例 $\alpha=\frac{FP}{TN+FP}$

假阴性率 假阴性样本占总阳性样本比例 $\beta=\frac{FN}{TP+FN}$

灵敏度 阳性样本正确识别出来的能力 $S_n=\frac{TP}{TP+FN}=1-\beta$

特异度 阴性样本正确识别出来的能力 $S_p=\frac{TN}{TN+FP}=1-\alpha$

2、Neyman-Pearson 决策

某些问题中一种错误比另一种错误重要，需在某个约束条件下最小化总风险

Neyman-Pearson 准则：严格限制较重要的一类错误概率为某常数而使另一类误判概率最小 → 常数

判决准则：$P_2(e)=\varepsilon_0$ 情况下使 $P_1(e)$ 最小

对于二分类问题，可能发生两种错误

① 模式 w_1 被误分到 w_2

② 模式 w_2 被误分到 w_1

设这两种错误的概率分别为 $P_1(e)$ 和 $P_2(e)$，有 $\Gamma_1,\Gamma_2,P(x|w_1),P(x|w_2)$ 对 $P_1(e),P_2(e)$ 产生影响

$$P_1(e)=\int_{\Gamma_2}P(x|w_1)dx,\quad P_2(e)=\int_{\Gamma_1}P(x|w_2)dx$$

其中 w_1 和 w_2 对应区域 Γ_1 和 Γ_2，比较平均错误率定义

$$P(e)=P_1(e)P(w_1)+P_2(e)P(w_2)=P(w_1)\int_{\Gamma_2}P(x|w_1)dx+P(w_2)\int_{\Gamma_1}P(x|w_2)dx$$

☆运用 Largrange 乘子法，取辅助函数 $L=P_1(e)+\lambda(P_2(e)-\varepsilon_0)$

$$L=P_1(e)+\lambda(P_2(e)-\varepsilon_0)=\int_{\Gamma_2}P(x|w_1)dx+\lambda[\int_{\Gamma_1}P(x|w_2)dx-\varepsilon_0]$$
$$=1-\int_{\Gamma_1}P(x|w_1)dx+\lambda\int_{\Gamma_1}P(x|w_2)dx-\lambda\varepsilon_0$$
$$=(1-\lambda\varepsilon_0)-\int_{\Gamma_1}[P(x|w_1)-\lambda P(x|w_2)]dx$$

设 Γ_1 和 Γ_2 的分界面为t，求使L最小的决策边界t，求L偏导并令其为0

$\begin{cases}\frac{\partial L}{\partial t}=0 \\ \frac{\partial L}{\partial \lambda}=0\end{cases}$ $\xrightarrow{决策边界上}$ $\begin{cases}\int_{\Gamma_1}P(x|w_2)dx=\varepsilon_0 & 约束 \\ \lambda=\frac{P(x|w_1)}{P(x|w_2)}\end{cases}$

得判决准则 ⇒令一类错误率为常数，使另一类错误率最小

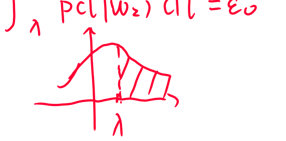
$\begin{cases}\frac{P(x|w_1)}{P(x|w_2)}>\lambda, & x\in w_1 \\ \frac{P(x|w_1)}{P(x|w_2)}<\lambda, & x\in w_2\end{cases}$ 由阈值决定

☆ $l(x)=\frac{P(x|w_1)}{P(x|w_2)}$

$\begin{cases} N-P & l(x)\gtrless\lambda\ (\int_{\Gamma_1}P(x|w_2)dx=\varepsilon_0)，则x\in\begin{cases}w_1\\w_2\end{cases} \\ min\ error & l(x)\gtrless\frac{P(w_2)}{P(w_1)}，则x\in\begin{cases}w_1\\w_2\end{cases} \\ min\ risk & l(x)\gtrless\frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}\frac{P(w_2)}{P(w_1)}，则x\in\begin{cases}w_1\\w_2\end{cases}\end{cases}$

$P_2(e)=\int_{\Gamma_1}P(x|w_2)dx$

$P_2(e)=\int_{\lambda}^{+\infty}p(l|w_2)dl=\varepsilon_0$



N-P决策下λ的计算

令 $p(l|x)$ 为似然比 $l(x)$ 在条件 $x\in w_2$ 下的概率密度，因 $l(x)>\lambda$ 时判决 $x\in w_1$，λ可用下式判定

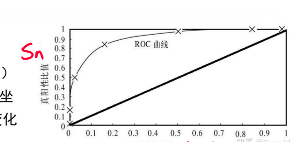
$$P_2(e)=\int_{\lambda}^{+\infty}p(l|w_2)dl=\varepsilon_0$$

因 $p(l|w_2)\geq 0$，$P_2(e)$ 关于λ单调，可以使用试探法计算 $P_2(e)=\varepsilon_0$ 时的λ，从中寻找使得 $P_1(e)$ 尽量小

3、ROC 曲线

ROC曲线 $S_n=\frac{TP}{TP+FN}$ $S_p=\frac{TN}{TN+FP}$

ROC曲线：如果把灵敏度即真阳性率（$1-P_1(e)$）作为纵坐标，假阳性率（$1-S_p=P_2(e)$）作为横坐标，可以获得随着决策面变化，两类错误率的变化情况。



似然函数

- 绘制方法：在类条件概率密度已知的情况下，可以通过求不同似然比阈值条件下的两类错误率（$P_1(e)=\int_{\Gamma_2}p(X|\omega_1)dX$ $P_2(e)=\int_{\Gamma_1}p(X|\omega_2)dX$），画出ROC曲线。
- 用途：① 比较不同分类判别方法的性能
  ② 评估特征与类别的相关性度量
- ROC曲线比较：可以用ROC曲线下的面积即AUC（Area under ROC curve）来定量衡量方法的性能。