§2.4 正态分布的统计决策

常为正态分布,数学易处理

※ 分类器决策面方程常涉及到 $\boxed{\text{类条件概率密度}}$ $p(x|\omega_i)$

1、正态分布

单变量：$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]$ ($\mu \to E[x]$, $\sigma \to$ std)

二项形式：$p(x)=(2\pi)^{-\frac{1}{2}}(\sigma^2)^{-\frac{1}{2}}\exp[-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)]$

多变量：$p(x)=(2\pi)^{-\frac{d}{2}}|\Sigma|^{-\frac{1}{2}}\exp[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)]$ 称x服从d维正态分布，记 $X\sim N_d(\mu,\Sigma)$

※式中 $x=[x_1,\cdots,x_d]^T$，$\mu=[\mu_1,\cdots,\mu_d]^T$，d维均值向量，$\mu=E\{x\}$；$\Sigma=E[(x-\mu)(x-\mu)^T]$，$d\times d$ 维协方差矩阵，对称且正定 $|\Sigma|>0$

性质：① 参数 μ 和 Σ 决定分布形状 $d+\frac{d(d+1)}{2}$ 个参数

② 等概率密度点轨迹为一超椭球面，μ 确定中心，Σ 确定大小

※ 马氏距离 $r^2=(x-\mu)^T\Sigma^{-1}(x-\mu)$，称x到$\mu$的马氏距离平方，等概率密度点轨迹是x到$\mu$马氏距离为常数的超椭球面，该超椭球体体积 $V=V_d|\Sigma|^{\frac{1}{2}}r^d$，$V_d$（d维单位超球体体积）$=\begin{cases}\frac{\pi^{\frac{d}{2}}}{(\frac{d}{2})!} & d\text{为偶数}\\ \frac{2^d\pi^{(d-1)/2}(\frac{d-1}{2})!}{d!} & d\text{为奇数}\end{cases}$

③ 正态分布中不相关 = 独立性 $\begin{cases}\text{不相关} & E[x_i,x_j]=E[x_i]E[x_j]\\ \text{独立性} & p(x_i,x_j)=p(x_i)p(x_j)\end{cases}$ Σ为对角阵　条件更强

④ 多元正态分布的边缘分布/条件分布具有正态性

⑤ 线性变换的正态性 $y=Ax$，$N(A\mu, A\Sigma A^T)$

⑥ 线性组合的正态性 $y=a^T\vec{x}$，$N(a^T\vec{\mu}, a^T\Sigma a)$

2、正态分布概率模型下的最小错误率贝叶斯决策

类条件概率密度为正态分布 $p(x|\omega_i)=\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_i|^{\frac{1}{2}}}\exp[-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)]$

有判别函数 $g_i(x)=\ln[p(x|\omega_i)p(\omega_i)]=\ln p(x|\omega_i)+\ln p(\omega_i)=-\frac{1}{2}((x-\mu_i)\Sigma_i^{-1}(x-\mu_i)^T)-\frac{d}{2}\ln 2\pi-\frac{1}{2}\ln|\Sigma_i|+\ln P(\omega_i)$

有决策面方程 $g_i(x)=g_j(x)$ 或 $g_i(x)-g_j(x)=0$ (i,j相邻) $\Rightarrow -\frac{1}{2}[(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)-(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)]-\frac{1}{2}\ln\frac{|\Sigma_i|}{|\Sigma_j|}+\ln\frac{P(\omega_i)}{P(\omega_j)}=0$

※有如下讨论

① $\Sigma_i=\sigma^2 I$，$g_i(x)=-\frac{1}{2}((x-\mu_i)\Sigma_i^{-1}(x-\mu_j)^T)-\frac{d}{2}\ln 2\pi-\frac{1}{2}\ln|\Sigma_i|+\ln P(\omega_i)$

各x_i统计量独立且方差相同 ⇒ 各样本落在以μ_i为中心同样大小的一些超球体中

此时 $|\Sigma_i|=\sigma^{2d}$，$\Sigma_i^{-1}=\frac{1}{\sigma^2}I$，得

$g_i(x)=-\frac{(x-\mu_i)(x-\mu_i)^T}{2\sigma^2}\underbrace{-\frac{d}{2}\ln 2\pi-\frac{1}{2}\ln\sigma^{2d}}_{\text{与类别无关}}+\ln P(\omega_i) \Longrightarrow g_i(x)=-\frac{\|x-\mu_i\|^2}{2\sigma^2}+\ln P(\omega_i)$

a. 若 $P(\omega_i)=P(\omega_j)$ 先验等

将待分类向量分到欧氏距离最小类中，$\min_{i=1,\cdots,c}\|x-\mu_i\|^2$，最小距离分类器/模板匹配

b. 若 $P(\omega_i)\neq P(\omega_j)$

将欧氏距离平方用σ^2规范化减去$\ln P(\omega_i)$再分类；若待分类模式距两类均值向量距离等，则将该模式分到先验大的那类

对 $g_i(x)=-\frac{\|x-\mu_i\|^2}{2\sigma^2}+\ln P(\omega_i)=-\frac{1}{2\sigma^2}(x^Tx-2\mu_i^Tx+\mu_i^T\mu_i)+\ln P(\omega_i)$ （x^Tx 与分类无关而忽略）

$\Rightarrow -\frac{1}{2\sigma^2}(-2\mu_i^Tx+\mu_i^T\mu_i)+\ln P(\omega_i)=W_i^Tx+W_{i0}$ ⇒ 线性判别函数，线性分类器

$W_i=\frac{1}{\sigma^2}\mu_i$，$W_{i0}=-\frac{1}{2\sigma^2}\mu_i^T\mu_i+\ln P(\omega_i)$

求决策面，$g_i(x)-g_j(x)=0$

$\Rightarrow \frac{1}{\sigma^2}(\mu_i^T-\mu_j^T)x-\frac{1}{2\sigma^2}(\mu_i^T\mu_i-\mu_j^T\mu_j)+\ln\frac{P(\omega_1)}{P(\omega_2)}=0$

$\|\mu_i-\mu_j\|^2=(\mu_i-\mu_j)^T(\mu_i-\mu_j)$　$\Rightarrow \frac{1}{\sigma^2}(\mu_i^T-\mu_j^T)[x-\frac{1}{2}(\mu_i+\mu_j)+\frac{\sigma^2(\mu_i-\mu_j)}{\|\mu_i-\mu_j\|^2}\ln\frac{P(\omega_1)}{P(\omega_2)}]=0$

$\Rightarrow W^T(x-x_0)=0$ 通过x_0正交于向量W的超平面

$\boxed{W=\mu_i-\mu_j}$（均值向量间连线），$x_0=\frac{1}{2}(\mu_i+\mu_j)-\frac{\sigma^2(\mu_i-\mu_j)}{\|\mu_i-\mu_j\|^2}\ln\frac{P(\omega_1)}{P(\omega_2)}$

a. $P(\omega_i)=P(\omega_j)$ 先验等

$x_0=\frac{1}{2}(\mu_i+\mu_j)$，超平面通过$\mu_i$与$\mu_j$连线中点$x_0$与连线正交

b. $P(\omega_i)\neq P(\omega_j)$ 先验不等

x_0不在中点，超平面向概率小一侧移动

② $\Sigma_i=\Sigma$

各类协方差矩阵相等 $\Sigma_1=\cdots=\Sigma_c=\Sigma$。几何上相当于各类样本集中于以该类均值$\mu_i$点为中心的同样大小和形状的超椭球体中。

判别函数 $g_i(x)=-\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i)+\ln P(\omega_i)$

若c类先验概率相等（即 $\ln P(\omega_1)=\ln P(\omega_2)=\cdots=\ln P(\omega_c)$），此时

$g_i(x)=(x-\mu_i)^T\Sigma^{-1}(x-\mu_i)=r^2$ — $\boxed{\text{马氏距离平方}}$ 考虑各种特性之间的联系　独立于测量尺度

将样本x归于到均值点马氏距离最近的类中

※ Σ为单位阵时，马氏距离退化为欧氏距离；Σ为对角阵时，马氏距离为正规化欧氏距离

求决策面，$g_i(x)-g_j(x)=0$

$\Rightarrow 2(\mu_i-\mu_j)\Sigma^{-1}x+(\mu_j^T\Sigma^{-1}\mu_j-\mu_i^T\Sigma^{-1}\mu_i)+(\ln P(\omega_i)-\ln P(\omega_j))=0$

$\Rightarrow 2(\mu_i-\mu_j)^T\Sigma^{-1}[x+\frac{(\mu_j^T\Sigma^{-1}\mu_j-\mu_i^T\Sigma^{-1}\mu_i)}{2(\mu_i-\mu_j)^T\Sigma^{-1}}+\frac{1}{(\mu_i-\mu_j)^T\Sigma^{-1}}\ln\frac{P(\omega_i)}{P(\omega_j)}]=0$

$\Rightarrow W^T(x-x_0)=0$ W与$x-x_0$正交，决策面过x_0

$W=\Sigma^{-1}(\mu_i-\mu_j)$（通常不在$\mu_i-\mu_j$方向上），$x_0=\frac{1}{2}(\mu_i+\mu_j)-\frac{(\mu_i-\mu_j)}{(\mu_i-\mu_j)^T\Sigma^{-1}(\mu_i-\mu_j)}\ln\frac{P(\omega_i)}{P(\omega_j)}$

a. 先验概率等，交点在均值向量连线中点

b. 先验概率不等，向小先验概率方向移动；相差较大时，判别边界不会落入球状分布中心点之间

(3) $\Sigma_i\neq\Sigma_j$，$i,j=1,2,\cdots,c$　$g_i(x)=-\frac{1}{2}((x-\mu_i)\Sigma_i^{-1}(x-\mu_i)^T)-\frac{d}{2}\ln 2\pi-\frac{1}{2}\ln|\Sigma_i|+\ln P(\omega_i)$

判别函数：$g_i(x)=-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)-\frac{1}{2}\ln|\Sigma_i|+\ln P(\omega_i)$

$=\boxed{x^TW_ix+w_i^Tx+w_{i0}}$　这是x的一个非线性二次形式

其中，$W_i=-\frac{1}{2}\Sigma_i^{-1}$ $(d\times d)$ 矩阵，$w_i=\Sigma_i^{-1}\mu_i$ (d维向量)，$w_{i0}=-\frac{1}{2}\mu_i^T\Sigma_i^{-1}\mu_i-\frac{1}{2}\ln|\Sigma_i|+\ln P(\omega_i)$

若决策域 R*i* 和 R*j* 毗邻，则决策面方程为：

$$g_i(x)-g_j(x)=0$$

$$x^T(W_i-W_j)x+(w_i-w_j)^Tx+w_{i0}-w_{j0}=0$$

其决策面是二次曲线(椭圆、双曲线、抛物线、一对直线)，Bayes分类器是二次曲线分类器。

§2.5 分类器错误率

1、最小错误率贝叶斯决策的错误率

$P(e)=P(\omega_1)\int_{\Gamma_2}p(\vec{x}|\omega_1)d\vec{x}+P(\omega_2)\int_{\Gamma_1}p(\vec{x}|\omega_2)d\vec{x}$

$=P(\omega_1)P_1(e)+P(\omega_2)P_2(e)$

最小错误率贝叶斯决策可保证决策的错误率在统计意义上是最小的，但分类器性能比较需知道大小？

✓ x为多维向量时，要进行多重积分，当概率密度表达式复杂时，难于计算，并且，积分范围中判决区域的不连续也导致直接的计算困难。

错误率计算途径：

(1) 理论计算

(2) 计算错误率上界

(3) 实验估计

只能在某些特殊情况下才能实现错误率的理论计算：

- 正态分布且等协方差阵（$\Sigma_i=\Sigma_j=\Sigma$）
- 随机变量各分量独立，且维数足够大

2、正态模式分类的错误率

见PPT

只要两类马氏距离足够大，误判概率可足够小，

$P(e)=\int_{\frac{\eta}{2}}^{\infty}\frac{1}{\sqrt{2\pi}}\exp[-\frac{y^2}{2}]dy$

3、独立随机变量错误率

$P(e)=P(\omega_i)\Phi(\frac{\theta_{ij}-\mu_i}{\sigma_i})+P(\omega_j)[1-\Phi(\frac{\theta_{ij}-\mu_j}{\sigma_j})]$

$\Phi(\mu)=\int_{-\infty}^{\mu}\frac{1}{\sqrt{2\pi}}\exp(-\frac{y^2}{2})dy$

§2.6 离散概率模型统计决策

此时 $p(x|\omega_j)$ 奇异化，积分形式 $\int p(x|\omega_j)dx \Rightarrow \sum_x p(x|\omega_j)$，概率密度函数 ⇒ 概率分布函数

独立的二值特征

考虑两类问题，其中特征向量的元素为二值的，并且条件独立。

令 $x=(x_1,...,x_d)'$，其中，x_i 可能为0或1，且：

$p_i=P[x_i=1|\omega_1]$　　$q_i=P[x_i=1|\omega_2]$

假设条件独立，可将x元素的概率写为 $P(x|\omega_i)$，即

$\begin{cases}P(x|\omega_1)=\prod_{i=1}^{d}p_i^{x_i}(1-p_i)^{1-x_i}\\ P(x|\omega_2)=\prod_{i=1}^{d}q_i^{x_i}(1-q_i)^{1-x_i}\end{cases}$ ⟶ $p_1=P[x_1=1|\omega_1]$，$p_2=P[x_2=1|\omega_1]$

$P((1,1)|\omega_1)=p_1p_2$

$P((1,0)|\omega_1)=p_1(1-p_2)$

$P((0,1)|\omega_1)=(1-p_1)p_2$

$P((0,0)|\omega_1)=(1-p_1)(1-p_2)$

那么似然比为

$\boxed{\frac{P(x|\omega_1)}{P(x|\omega_2)}=\prod_{i=1}^{d}(\frac{p_i}{q_i})^{x_i}(\frac{1-p_i}{1-q_i})^{1-x_i}}$

独立的二值特征

$\boxed{\frac{P(x|\omega_1)}{P(x|\omega_2)}=\prod_{i=1}^{d}(\frac{p_i}{q_i})^{x_i}(\frac{1-p_i}{1-q_i})^{1-x_i}}$

基于最小错误率贝叶斯准则的判决函数 $g(x)=\ln\frac{p(x|\omega_1)}{p(x|\omega_2)}+\ln\frac{P(\omega_1)}{P(\omega_2)}$

$g(x)=\sum_{i=1}^{d}[x_i\ln\frac{p_i}{q_i}+(1-x_i)\ln\frac{1-p_i}{1-q_i}]+\ln\frac{P(\omega_1)}{P(\omega_2)}$

注意判决函数对x_i是线性的，可改写为

$g(x)=\sum_{i=1}^{d}w_ix_i+w_0$　　其中，$w_i=\ln\frac{p_i(1-q_i)}{q_i(1-p_i)}$　$i=1,...,d$　$w_0=\sum_{i=1}^{d}\ln\frac{1-p_i}{1-q_i}+\ln\frac{P(\omega_1)}{P(\omega_2)}$

若 $g(x)>0$ 判别为 ω_1，否则为 ω_2。

- $g(x)$可以看作是x的各分量的加权组合
- 特征独立的条件产生线性分类器，而如果特征不独立将产生复杂的分类器