是一个比分类器设计问题更难的一般性问题①，我们试图通过解决这个更难的一般问题来解决分类问题显然是不合理的。所以，在一些实际工程问题中，下面几章介绍的直接设计分类器的方法经常得到更广泛的应用。

习　　题

3.1　设总体分布密度为 $N(\mu,1)$，$-\infty<\mu<+\infty$，并设 $\mathscr{X}=\{x_1,x_2,\cdots,x_N\}$，分别用最大似然估计和贝叶斯估计计算 $\hat{\mu}$。已知 μ 的先验分布 $p(\mu)\sim N(0,1)$。

3.2　设 $\mathscr{X}=\{x_1,x_2,\cdots,x_N\}$为来自总体 $N(\mu,\sigma^2)$的样本集，求 μ,σ^2 的最大似然估计 $\hat{\mu},\hat{\sigma}^2$。

3.3　设 $\mathscr{X}=\{x_1,x_2,\cdots,x_N\}$为来自点二项分布的样本集，即

$$f(x,P)=P^xQ^{(1-x)},x=0,1,0\leqslant P\leqslant 1,Q=1-P$$

试求参数 P 的最大似然估计量 $\hat{P}$。

3.4　假定损失函数为二次函数 $\lambda(\hat{P},P)=(\hat{P}-P)^2$，以及 P 的先验密度为均匀分布即 $f(P)=1,0\leqslant P\leqslant 1$。在这样的假设条件下，求 3.3 题的贝叶斯估计量 $\hat{P}$。

3.5　用贝叶斯学习的性质，说明当样本数 N 趋于无穷时，最大似然估计将等价于贝叶斯估计。

3.6　对正态分布的期望 μ 的贝叶斯学习所得到的分布 $p(x|\mathscr{X})$的方差为 $\sigma^2+\sigma_N^2$。指出

(1)该方差为什么大于 σ^2；

(2)当 $N\to\infty$时，$\sigma^2+\sigma_N^2$ 有什么变化；

(3)利用贝叶斯学习的性质分析(2)中的结果。

3.7　设 $\mathscr{X}=\{x_1,x_2,\cdots,x_N\}$是来自 $p(x|\theta)$的随机样本，其中 $0\leqslant x\leqslant\theta$ 时，$p(x|\theta)=\frac{1}{\theta}$，否则为 0。证明 θ 的最大似然估计是$\max\limits_k x_k$。

3.8　利用矩阵恒等式

$$(A^{-1}+B^{-1})^{-1}=A(A+B)^{-1}B=B(A+B)^{-1}A$$

对于多元正态分布，写出与式(3-53)和式(3-54)相应的结果。

3.9　设 $\mathscr{X}=\{x_1,x_2,\cdots,x_N\}$来自数学期望为 μ 的总体，证明统计量

$$\varphi_1\{x_1,x_2,\cdots,x_N\}=x_i,\quad i=1,2,\cdots,N$$

及

$$\varphi_2\{x_1,\cdots,x_N\}=\sum_{i=1}^{N}\alpha_ix_i\quad 其中\sum_{i=1}^{N}\alpha_i=1$$

都是 μ 的无偏估计。

3.10　证明对正态总体的期望 μ 和方差 σ^2 的最大似然估计 $\hat{\mu}$ 是无偏的，而 $\hat{\sigma}^2$ 是有偏的。

3.11　利用待定系数法，推导式(3-53)和式(3-54)。

---

① 参见 Vladimir Vapnik，Estimation of Dependences Based on Empirical Data. Springer-Verlag NY，1982 和参考书目[8]。

3.12 推导式(3-70)。

3.13 用 Lagrange 乘子法推导式(3-74),条件为式(3-72)和式(3-73)。

3.14 对于多元正态分布 $p(x) \sim N(\mu, \Sigma)$,若 Σ 为已知,通过样本集 $\mathscr{X} = \{x_1, x_2, \cdots, x_N\}$ 求 μ 的最大似然估计量 $\hat{\mu}$。

3.15 设 $p(x) \sim N(\mu, \sigma^2)$,窗函数 $\varphi(x) \sim N(0,1)$,指出 Parzen 窗估计

$$\hat{p}_N(x) = \frac{1}{Nh_N} \sum_{i=1}^{N} \varphi\left(\frac{x - x_i}{h_N}\right)$$

对于小的 h_N,有如下性质:

(1) $E[\hat{p}_N(x)] \sim N(\mu, \sigma^2 + h_N^2)$

(2) $\mathrm{Var}[\hat{p}_N(x)] = \frac{1}{Nh_N 2\sqrt{\pi}} p(x)$

3.16 给出 Parzen 窗估计的程序框图,并编写程序。

3.17 假定式(3-127)的 ε 的密度函数可以由一未知期望值 ε 和已知方差 $\hat{\varepsilon}(1-\hat{\varepsilon})/N$ 的正态分布来近似,当 $N=250$ 及 $\hat{\varepsilon}=0.2$ 时,求对于置信度 $\gamma=0.95$ 的 ε 的置信区间,并将此结果与图 3.9 得出的结果作比较。