1. Describe the structure of a pattern classification system and give detailed information about each module.
2. Vessel diseases are a growing problem in the western world. Now, there is a software that can classify a diseased person as actually diseased with 99% reliability. However,it may happen in 2% of the cases that a healthy person is mistakenly classified as diseased. A statistical analysis shows that the disease is apparent in one out of 100 patients. What is the probability that a patient is actually diseased if the system classifies a disease?
3. 分别写出在以下两种情况

$$P(x|w_1) = P(x|w_2)$$
$$P(w_1) = P(w_2)$$

   下的最小错误率贝叶斯决策规则。

4. 若$\lambda_{11} = \lambda_{22} = 0$，$\lambda_{12} = \lambda_{21}$ ，证明此时最小最大决策面是来自两类的错误率相等。
5. 二维正态分布，$\mu_1 = (-1,0)^T$，$\mu_2 = (1,0)^T$，$\Sigma_1 = \Sigma_2 = I$，$P(\omega_1) = P(\omega_2)$。试写出对数似然比决策规则。
6.

(1) 假设某部位的细胞识别中正常（ϖ_1）和异常（ϖ_2）两类的先验概率分别为正常状态：$P(\varpi_1) = 0.8$；异常状态：$P(\varpi_2) = 0.2$。现有一个待识细胞，其观察值为x，从类条件概率密度分布曲线上查得：$P(x|\varpi_1) = 0.3$，$P(x|\varpi_2) = 0.6$，用最小错误率贝叶斯规则对该细胞x进行分类。

(2) 对 (1) 中的条件，利用下面的决策表，按最小风险贝叶斯决策进行分类。

状态 损失 决策	ϖ_1	ϖ_2
a_1	0	7
a_2	1	0

7. 设在一维特征空间中两类样本服从正态分布，$\sigma_1 = \sigma_2 = 2$，μ1=0，μ2=3，两类先验概率之比$P(\omega_1)/P(\omega_2) = e$，试求按基于最小错误率贝叶斯决策原则的决策分界面的 x 值。

8. 设总体分布密度为$N(\mu, 1)$，$-\infty < \mu < +\infty$，并设$X = \{x_1, x_2, \ldots, x_N\}$，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 $p(\mu) \sim N(0,1)$。

1. 焦点知识

信息获取-》预处理模块-〉特征提取模块-》决策

系统知识

焦点知识：现场获得的知识
系统知识：模式识别系统经过一个稳定的学习获得一个稳定的知识

2

$$P(A|B) = \frac{0.99 \cdot 0.01}{0.0297} = \frac{0.0099}{0.0297}$$

3

1) 对比先验概率

2) 对比似然比

4

在最小最大决策面下，我们的目标是最大化正确分类的概率，或者等价地，最小化最大的可能损失。

5

6

1) w1=0.24>w2=0.12
2)分为第二类

7

7. $P(w_i|x) = \frac{P(x|w_i)P(w_i)}{P(x)}$ => 决策面 $P(x|w_1)\overset{e}{\cancel{P(w_1)}} = P(x|w_2)\cancel{P(w_2)}$

$P(x|w_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}$

因此 $e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \cdot e = e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$ => 同时对数

=> $-\frac{x^2}{8} + 1 = -\frac{(x-3)^2}{8}$ => 解得 $x = \frac{17}{6}$

8

9. 设 $X=\{x_1,x_2,...,x_N\}$为来自二项分布的样本集，即 $f(x,P)=P^xQ^{(1-x)},x=0,1$，$0\le P\le 1$，$Q=1-P$，试求参数 P 的最大似然估计。

10. 考虑一维正态分布的参数估计。设样本（一维）x_1，x_2，...x_N都是由独立的抽样试验采集的，且密度函数服从正态分布，其均值μ与方差σ^2未知。求均值和方差的最大似然估计。

11. 设一维样本集 X=$\{x_1$，x_2，...$x_N\}$是取自正态分布 $N(\mu$，$\sigma^2)$的样本集，其中均值μ为未知的参数，方差σ^2已知。未知参数μ是随机参数，它有先验分布 $N(\mu_0$，$\sigma_0^2)$的，μ_0、σ_0^2已知，求μ的贝叶斯估计$\hat{\mu}$。

12. 线性分类器的分界面是超平面，线性分类器设计步骤是什么？

13. Fisher 线性判别函数是研究这类判别函数中最有影响的方法之一,请简述它的准则.

14. 感知器的准则函数是什么？它通过什么方法得到最优解？

15. 设有一维空间二次判别函数

$$g(x)=5+7x+9x^2$$

（1） 试映射成广义齐次线性判别函数；

（2） 总结把高次函数映射成齐次线性函数的方法。

16. 对于二维线性判别函数 $g(x)=x_1+2x_2-2$

（1） 将判别函数写成 $g(x)=w^Tx+w_0$的形式，并画出 $g(x)=0$ 的几何图形；

（2） 映射成广义齐次线性函数 $g(x)=a^Ty$；

（3） 指出上述 X 空间实际是 Y 空间的一个子空间，且$a^Ty=0$ 对于 X 子空间的划分和原空间中$w^T+w_0=0$ 对原 X 空间的划分相同，并在图上表示出来。

17. 指出在 Fisher 线性判别中，w 的比例因子对 Fisher 判别结果无影响。 $J(W)=\frac{(W^Tm_1-W^Tm_2)^2}{W^TS_1W+W^TS_2W}$

18. 证明在正态等方差条件下，Fisher 线性判别等价于贝叶斯判别。 $J(k\cdot W)=\frac{(k\cdot W^Tm_1-k\cdot W^Tm_2)^2}{k^2\cdot W^TS_1W+k^2\cdot W^TS_2W}$

19. 考虑准则函数

$$J(a)=\sum_{y\in\mathcal{Y}(a)}(a^Ty-b)^2$$

其中$\mathcal{Y}(a)$是使$a^Ty\le b$的样本集合。设y_1是$\mathcal{Y}(a)$中的唯一样本，则 J(a)的梯度为

12. 给定有标注的类别样本集、
建立准则函数J
使用最优化技术取得J的极致解

13

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

14

$$\mathbf{g(x)} = \mathbf{w}^T\mathbf{x} + w_0 = \mathbf{a}^T\mathbf{y}$$

梯度下降的方法

15

解：（1）设$y = [y_1, y_2, y_3]^T = [1, x, x^2]^T$，$a = [5, 7, 9]^T$，则广义齐次线性判别函数为: $g(x) = a^T y$

（2）对于n次函数$g(x) = c_0 + c_1 x + c_2 x^2 + ... + c_n x^n$，令$y = [y_1, y_2, ..., y_{n+1}]^T = [1, x, ..., x^n]^T$，$a = [c_0, c_1, ..., c_n]^T$，则$g(x) = a^T y$。

16

解：（1）$w = [1, 2]^T, x = [x_1, x_2]^T, w_0 = -2$，则$g(x) = w^T x + w_0$，$g(x) = 0$的图形如下图2:

（2）$y = [y_1, y_2, y_3]^T = [1, x_1, x_2]^T$，$a = [-2, 1, 2]^T$，则$g(x) = a^T y$。

（3）$y_1 = 1, y_2 = x_1, y_3 = x_2$，在所以所有的样本在Y空间中的一个平面$y_1 = 1$上。

17 记住Fisher两种准则的写法即可

18

$\nabla J(a) = 2(a_k^T y_1 - b)y_1$，二阶偏导数矩阵 $D = 2y_1 y_1^T$。据此证明，若最优步长选择为$\rho_k = \frac{||\nabla J(a)||^2}{\nabla J^T(a) D \nabla J(a)}$时，梯度下降法的迭代公式为：

可化简rou=1/2lly1ll^2

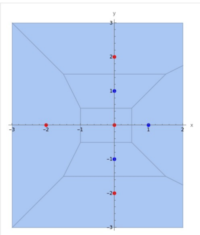
$$a_{k+1} = a_k + \frac{b - a_k^T y_1}{||y_1||^2} y_1$$

20. 在多类问题中，如果一组样本可被一线性机全部正确分类，则称这组样本是线性可分的。对任意w_i类，如果能用一超平面把w_i类的样本同其他样本分开来，则称总体线性可分。举例说明，总体线性可分必定线性可分，但反之不然。

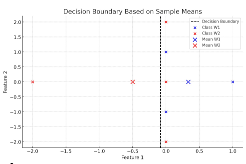
对于2类的问题，可以设X1和X2类，由x=c区分，若他是总体可分问题，则说明，对于X1类，必定有x=c1可以区分X1类与其他类，同理有X2类有x=c2可以区分其他类，此时已经线性可分

21. 设两类样本的类内离散矩阵分别为 $S_1 = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$，$S_2 = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}$ 均值向量 $m_1 = (2,0)^t, m_2 = (2,2)^t$ 试用 Fisher 准则求其决策面方程.

W*=Sw^-1*(m1-m2)=(0,-1)　　因此：X2=-1



22. 有七个二维向量：$x_1^T = (1,0)$、$x_2^T = (0,1)$、$x_3^T = (0,-1)$、$x_4^T = (0,0)$、$x_5^T = (0,2)$、$x_6^T = (0,-2)$、$x_7^T = (-2,0)$，假定前三个为w_1类，后四个为w_2类。



（1）画出最近邻法决策面；

（2）求样本均值m_1、m_2。若按离样本均值距离的大小进行分类，试画出决策面。

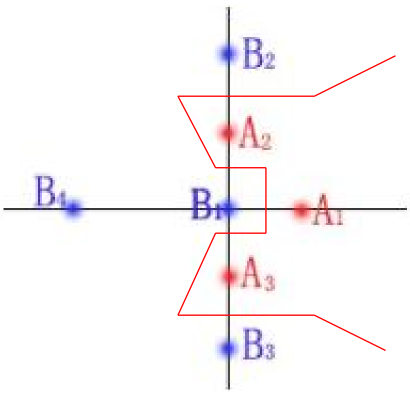
23. 设在一个二维空间，A 类有三个训练样本，图中用红点表示，B 类四个样本，图中用蓝点表示。

试问：

（1）按近邻法分类，这两类最多有多少个分界面　12个

（2）画出实际用到的分界面　　A3-B2，A2-B3，A1-B4的没有用到

（3） A1 与 B4 之间的分界面有没有用到？ 实际并没有用到



19

证明： y_1是$\mathscr{Y}(a)$中的唯一样本，则准则函数为$J(a)=\sum_{y\in\mathscr{Y}(a)}(a^Ty-b)^2=(a^Ty_1-b)^2$，所以$\nabla J(a)=2(a^Ty_1-b)y_1$，二阶偏导数矩阵为$D=2y_1y_1^T$。

梯度下降的迭代公式为：$a_{k+1}=a_k-\rho_k\nabla J(a_k)$，$\rho_k=\dfrac{4(a_k^Ty_1-b)^2||y_1||^2}{8(a_k^Ty_1-b)^2y_1^Ty_1y_1^Ty_1}=\dfrac{1}{2||y_1||^2}$，将$\rho_k$代入梯度下降的迭代公式：$a_{k+1}=a_k+\dfrac{b-a_k^Ty_1}{||y_1||^2}y_1$

20

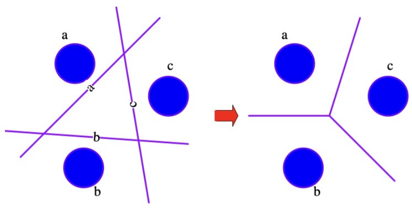


图 3: 总体线性可分必定线性可分

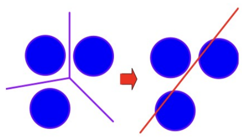


图 4: 线性可分未必总体线性可分

21