中国科学院大学

试 题 专 用 纸

课程编号：091M4042H
课程名称：模式识别与机器学习
任课教师：黄庆明、山世光、常虹、兰艳艳

注意事项：
1.考试时间为 120 分钟，考试方式 闭 卷；
2.全部答案写在答题纸上；
3.考试结束后，请将本试卷和答题纸、草稿纸一并交回。

以一条直线对模式进行分类，该直线的方程被描述成线性判别函数.

一般形式
$g(x)=w^T x+b$
权向量　阈值权

高维投影到低维线性空间

1. (8 分) 试阐述线性判别函数的基本概念，并说明既然有线性判别函数，为什么还需要非线性判别函数？假设有$\boxed{\text{两类模式}}$，每类包括 5 个 3 维不同的模式，且良好分布。如果它们是线性可分的，问权向量至少需要几个系数分量？假如要建立二次的多项式判别函数，又至少需要几个系数分量？（设模式的良好分布不因模式变化而改变）

4

2. (8 分) 简述偏差方差分解及其推导过程，并说明偏差、方差和噪声三部分的内在含义。

3. (8 分) 试描述用 EM 算法求解高斯混合模型的思想和过程，并分析 k-means 和高斯混合模型在求解聚类问题中的异同。

4. (10 分) 用下列势函数

$$K(x,x_k)=e^{-\|x-x_k\|^2}$$

求解以下模式的分类问题

ω_1: $\{(0\ \ 1)^T,\ (0\ \ -1)^T\}$

ω_2: $\{(1\ \ 0)^T,\ (-1\ \ 0)^T\}$

5. (10 分) 试述 K-L 变换的基本原理，并将如下两类样本集的特征维数降到一维，同时画出样本在该空间中的位置。

ω_1: $\{(-5\ \ -5)^T,\ (-5\ \ -4)^T,\ (-4\ \ -5)^T,\ (-5\ \ -6)^T,\ (-6\ \ -5)^T\}$

ω_2: $\{(5\ \ 5)^T,\ (5\ \ 6)^T,\ (6\ \ 5)^T,\ (5\ \ 4)^T,\ (4\ \ 5)^T\}$,

其中假设其先验概率相等，即 $P(\omega_1)=P(\omega_2)=0.5$。

6. (10 分) 详细描述 AdaBoost 算法，并解释为什么 AdaBoost 经常可以在训练误差为 0 后继续训练还可能带来测试误差的继续下降。

当AdaBoost的训练误差为0时 Adaboost会继续增大分类间距，提高模型泛化能力 降低模型误差

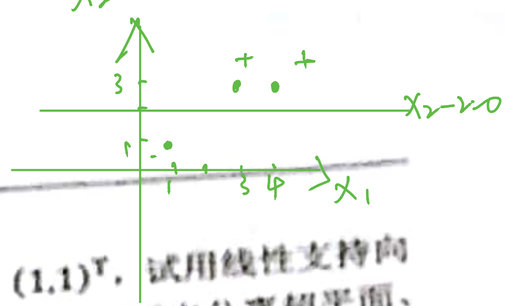
7. (10 分) 描述感知机（Perceptron）模型，并给出其权值学习算法。在此基础上，以仅有一个隐含层的三层神经网络为例，形式化描述 Back-Propagation（BP）算法中是如何对隐层神经元与输出层神经元之间的连接权值进行调整的。

SVM: $\max\limits_{\alpha} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j (x_i)^T x_j$

s.t. $\alpha_i \geq 0 \quad i=1\cdots N, \quad \sum_{i=1}^{N}\alpha_i y_i = 0$

$w^* = \sum_{i=1}^{N}\alpha_i^* y_i x_i$

$b^* = y_j - \sum_{i=1}^{N}\alpha_i^* y_i (x_i)^T x_j$



8. （12 分）已知正例点 $x_1=(3,3)^T$, $x_2=(4,3)^T$，负例点 $x_3=(1,1)^T$，试用线性支持向量机的对偶算法求最大间隔分离超平面和分类决策函数，并在图中画出分离超平面、间隔边界及支持向量。

(1,1)　(3,3)　(4,3)　$(0\ 1)\begin{pmatrix}x_1\\x_2\end{pmatrix}-2=0$　$x_2-2=0$　$w=(0,1)^T$　$b=-2$

$x_2-1=0$　$x_2-3=0$

9. （12 分）假定对一类特定人群进行某种疾病检查，正常人以 ω_1 类代表，患病者以 ω_2 类代表。设被检查的人中正常者和患病者的先验概率分别为

正常人：$P(\omega_1)=0.9$

患病者：$P(\omega_2)=0.1$

现有一被检查者，其观察值为 x，从类条件概率密度分布曲线上查得

$P(x|\omega_1)=0.2$, $P(x|\omega_2)=0.4$

同时已知风险损失函数为

$$\begin{pmatrix}\lambda_{11} & \lambda_{12}\\ \lambda_{21} & \lambda_{22}\end{pmatrix}=\begin{pmatrix}0 & 6\\ 1 & 0\end{pmatrix}$$

其中 λ_{ij} 表示将本应属于第 j 类的模式判为属于第 i 类所带来的风险损失。试对该被检查者用以下两种方法进行分类：

（1）　基于最小错误率的贝叶斯决策，并写出其判别函数和决策面方程；

（2）　基于最小风险的贝叶斯决策，并写出其判别函数和决策面方程。

10. （12 分）假设有 3 个盒子，每个盒子里都装有红、白两种颜色的球。按照下面的方法抽球，产生一个球的颜色的观测序列：开始，以概率 π 随机选取 1 个盒子，从这个盒子里以概率 B 随机抽出 1 个球，记录其颜色后，放回；然后，从当前盒子以概率 A 随机转移到下一个盒子，再从这个盒子里以概率 B 随机抽出一个球，记录其颜色，放回；如此重复进行 3 次，得到一个球的颜色观测序列：O =（红，白，红）。请计算生成该序列的概率 $P(O|\{A,B,\pi\})$。

提示：假设状态集合是{盒子 1，盒子 2，盒子 3}，观测的集合是{红，白}，本题中已知状态转移概率分布、观测概率分布和初始概率分布分别为：

$$A=\begin{array}{c|c} & \text{盒子1}\quad\text{盒子2}\quad\text{盒子3}\\ \begin{array}{c}\text{盒子1}\\ \text{盒子2}\\ \text{盒子3}\end{array} & \begin{bmatrix}0.5 & 0.2 & 0.3\\ 0.3 & 0.5 & 0.2\\ 0.2 & 0.3 & 0.5\end{bmatrix}\end{array},\quad B=\begin{array}{c|c} & \text{红}\quad\text{白}\\ \begin{array}{c}\text{盒子1}\\ \text{盒子2}\\ \text{盒子3}\end{array} & \begin{bmatrix}0.5 & 0.5\\ 0.4 & 0.6\\ 0.7 & 0.3\end{bmatrix}\end{array},\quad \pi=[0,2,0.4,0.4]^T。$$

中国科学院大学

试 题 专 用 纸

课程编号：091M4042H

课程名称：模式识别与机器学习

任课教师：黄庆明、山世光、兰艳艳、郭嘉丰

注意事项：

1.考试时间为 120 分钟，考试方式 闭 卷；

2.全部答案写在答题纸上；

3.考试结束后，请将本试卷和答题纸、草稿纸一并交回。

1. （6 分）简述模式的概念和它的直观特性，并简要说明模式分类有哪几种主要方法。

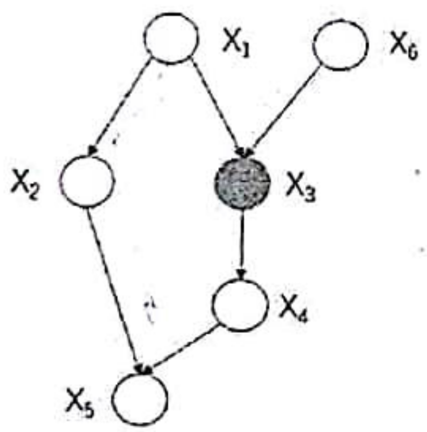
2. （8分）假设某研究者在ImageNet数据上使用线性支持向量机（Linear SVM）来做文本分类的任务，请说明在如下情况下分别如何操作才能得到更好的结果，并说明原因。

（1）训练误差5%，验证误差10%，测试误差10%。

（2）训练误差1%，验证误差10%，测试误差10%。

（3）训练误差1%，验证误差3%，测试误差10%。

3. （8分）给定如下概率图模型，其中变量X_3为已观测变量，请问变量X_4和X_6是否独立？并用概率推导证明之。



4. （10 分）（1）随机猜测作为一个分类算法是否一定比 SVM 差？借此阐述你对“No Free Lunch Theorem”的理解。（2）举例阐述你对“Occam's razor”的理解。

5. （10 分）详细描述 AdaBoost 的原理并给出算法，并解释为什么 AdaBoost 经常可以在训练误差为 0 后继续训练还可能带来测试误差的继续下降。

6. （10 分）用感知器算法求下列模式分类的解向量（取 w(1) 为零向量）

ω_1: $\{(0\ 0\ 0)^T,\ (1\ 0\ 0)^T,\ (1\ 0\ 1)^T,\ (1\ 1\ 0)^T\}$

ω_2: $\{(0\ 0\ 1)^T,\ (0\ 1\ 1)^T,\ (0\ 1\ 0)^T,\ (1\ 1\ 1)^T\}$

7. （12 分）设以下模式类别具有正态概率密度函数：

　　ω_1: $\{(0\ 0\ 0)^T,\ (1\ 0\ 0)^T,\ (1\ 0\ 1)^T,\ (1\ 1\ 0)^T\}$

　　ω_2: $\{(0\ 1\ 0)^T,\ (0\ 1\ 1)^T,\ (0\ 0\ 1)^T,\ (1\ 1\ 1)^T\}$

若 $P(\omega_1)=P(\omega_2)=0.5$，求这两类模式之间的贝叶斯判别界面的方程式。

8. （12 分）假设有如下线性回归问题，

$$\min_{\beta}(y-X\beta)^2+\lambda\|\beta\|_2^2$$

其中y和β是n维向量，X是一个$m\times n$的矩阵。

该线性回归问题的参数估计可看作一个后验分布的均值，其先验为高斯分布$\beta\sim N(0,\tau I)$，样本产生自高斯分布$y\sim N(X\beta,\sigma^2 I)$，其中$I$为单位矩阵，试推导调控系数$\lambda$与方差$\tau$和$\sigma^2$的关系。

9. （12 分）给定有标记样本集$D_l=\{(x_1,y_1),(x_2,y_2),\dots,(x_l,y_l)\}$和未标记样本$D_u=\{(x_{l+1},y_{l+1}),(x_{l+2},y_{l+2}),\dots,(x_{l+u},y_{l+u})\},l\ll u,l+u=m$，假设所有样本独立同分布，且都是由同一个包含 N 个混合成分的高斯混合模型$\{(\alpha_i,\mu_i,\Sigma_i)|1\le i\le N\}$产生，每个高斯混合成分对应一个类别，请写出极大似然估计的目标函数（对数似然函数），以及用 EM 算法求解参数的迭代更新式。

10. （12 分）假定对一类特定人群进行某种疾病检查，正常人以 ω_1类代表，患病者以 ω_2类代表。设被检查的人中正常者和患病者的先验概率分别为

　　正常人：$P(\omega_1)=0.9$

　　患病者：$P(\omega_2)=0.1$

现有一被检查者，其观察值为 x，从类条件概率密度分布曲线上查得

　　$P(x|\omega_1)=0.2$，$P(x|\omega_2)=0.4$

同时已知风险损失函数为

$$\begin{pmatrix}\lambda_{11} & \lambda_{12}\\ \lambda_{21} & \lambda_{22}\end{pmatrix}=\begin{pmatrix}0 & 6\\ 1 & 0\end{pmatrix}$$

其中 λ_{ij}表示将本应属于第 j 类的模式判为属于第 i 类所带来的风险损失。试对该被检查者用以下两种方法进行分类：

（1）　基于最小错误率的贝叶斯决策，并写出其判别函数和决策面方程；

（2）　基于最小风险的贝叶斯决策，并写出其判别函数和决策面方程。

2016-2017

1. 模式是抽取自物体的信息集合，既包含空间部分，又包含时间部分.

直观特性：可观察性，可区分性，相似性

主要方法：监督学习：概念驱动，归纳假说

非监督学习：数据驱动，演绎假说

2、(1) 欠拟合，换用复杂度更高的模型

(2) 过拟合，换用复杂度更低的模型

(3) 测试数据与训练数据不是独立同分布的，更换测试数据集

4. (1) 不一定，在无先验知识的情况下，无法断言一个模型比另一个更好.

对特定的问题，为了获得更好的效果需要使用更复杂的模型.

(2) 训练数据来自添加高斯噪声的 $y=\sin x \ (x\in[0,2\pi])$. 训练

使用不同的多项式函数拟合，三次的效果最佳. 在同等错误

率的条件下，简单模型具有更小的方差，更好的泛化能力.

5. $\{x_i, y_i\}_{i=1}^{n}$

$D(i)=\frac{1}{n}$

for 1 to T:

训练弱分类器 h_i

$D_{(i)}^{t+1} = D_{(i)}^{t} \cdot e^{-\alpha_i y_i h_i(x_i)}$

通过弱分类器的组合，得到强分类器.

每次训练弱分类器后，对分类错误的样本，增加权重使得后面的分类器更加"关注"该样本，以达到提升分类效果的目的.

其中 $\alpha_i = \frac{1}{2}\ln\frac{1-\varepsilon}{\varepsilon}$， $\varepsilon = P(h(x)\neq y) < 0.5$

$H_{final} = \sum \alpha_i h_i(x)$

当训练误差为零后，AdaBoost 会继续增大分类间距，提升模型的泛化能力，减少测试误差.

6. $W(1)=(0,0,0,0)$，对 ω_2 的样本取相反数

增广数据：$\omega_1: \{(0,0,0,1)^T, (1,0,0,1)^T, (1,0,1,1)^T, (1,1,0,1)^T\}$

$\omega_2: \{(-0,-0,-1,-1)^T, (-0,-1,-1,-1)^T, (-0,-1,-0,-1)^T, (-1,-1,-1,-1)^T\}$

更新规则：$w(k+1) = \begin{cases} w(k) & w(k)^T x(k) > 0 \\ w(k)+x(k+1) & 其它 \end{cases}$ （以下省略不更新的步骤）

$x(1)=(0,0,0,1)^T$，$w(1)^T x(1)=0$ 更新：$w(2)=(0,0,0,1)^T$

$x(2)=(0,0,-1,-1)^T$，$w(2)^T x(2)=-1$ 更新：$w(3)=(0,0,-1,0)^T$

$x(3)=(0,-1,0,-1)^T$，$w(3)^T x(3)=0$ 更新：$w(4)=(0,-1,-1,-1)^T$

$x(4)=(0,0,0,1)^T$，$w(4)^T x(4)=-1$，更新：$w(5)=(0,-1,-1,0)^T$

$x(5)=(1,0,0,1)^T$，$w(5)^T x(5)=0$，更新：$w(6)=(1,-1,-1,1)^T$

$x(6)=(0,0,-1,-1)^T$，$w(6)^T x(6)=0$，更新：$w(7)=(1,-1,-2,0)^T$

$x(7)=(0,0,0,1)^T$，$w(7)^T x(7)=0$，更新：$w(8)=(1,-1,-2,1)^T$

$x(8)=(1,0,1,1)^T$，$w(8)^T x(8)=0$，更新：$w(9)=(2,-1,-1,2)^T$

$x(9)=(0,0,-1,-1)^T$，$w(9)^T x(9)=-1$，更新：$w(10)=(2,-1,-2,1)^T$

$x(10)=(0,-1,0,-1)^T$，$w(10)^T x(10)=0$，更新：$w(11)=(2,-2,-2,0)^T$

$x(11)=(0,0,0,1)^T$，$w(11)^T x(11)=0$，更新：$w(12)=(2,-2,-2,1)^T$

$w=w(12)=(2,-2,-2,1)$

7. $p(\omega_1|x)=\dfrac{p(x|\omega_1)p(\omega_1)}{p(x)}$

$p(\omega_2|x)=\dfrac{p(x|\omega_2)p(\omega_2)}{p(x)}$

$p(\omega_1|x) > p(\omega_2|x)$

$p(x|\omega_1)p(\omega_1) > p(x|\omega_2)p(\omega_2)$

$\dfrac{p(x|\omega_1)}{p(x|\omega_2)} - \dfrac{p(\omega_2)}{p(\omega_1)} > 0$

$d(x)=\dfrac{p(x|\omega_1)}{p(x|\omega_2)} - \dfrac{p(\omega_2)}{p(\omega_1)} = \dfrac{p(x|\omega_1)}{p(x|\omega_2)} - 1 = 0$

假设给定ω的情况下，x的各分量相互独立。

$p(x|\omega)=p(x_1,x_2,x_3|\omega)=p(x_1|\omega)p(x_2|\omega)p(x_3|\omega)$

$dx=\dfrac{\prod_{i=1}^{3} p(x_i|\omega_1)}{\prod_{i=1}^{3} p(x_i|\omega_2)} - 1 = 0$

$$p(\vec{y})$$

对数似然：$\log p(\vec{\beta}|\vec{y}) = \log p(\vec{y}|\vec{\beta},x,\sigma) + \log p(\vec{\beta}|\tau) - \log p(\vec{y})$

$$= \log \prod_{i=1}^{n} p(y_i|\vec{\beta},\vec{x}_i,\sigma) + \log\frac{1}{\sqrt{2\pi|\tau I|}}\cdot e^{-\frac{1}{2}(\vec{\beta}-\vec{0})^T(\tau I)^{-1}(\vec{\beta}-\vec{0})} - \log p(\vec{y})$$

$$= \sum_{i=1}^{n}\log\frac{1}{\sqrt{2\pi|\sigma I|}} e^{-\frac{1}{2}(y_i-\vec{x}_i^T\vec{\beta})(\sigma I)^{-1}(y_i-\vec{x}_i^T\vec{\beta})} + \left(-\frac{1}{2}\vec{\beta}^T(\tau I)^{-1}\vec{\beta}\right) + \text{constant}$$

$$= \sum_{i=1}^{n} -\frac{1}{2}(y_i-\vec{x}_i^T\vec{\beta})^T(\sigma I)^{-1}(y_i-\vec{x}_i^T\vec{\beta}) - \frac{1}{2}\vec{\beta}^T(\tau I)^{-1}\vec{\beta} + \text{constant}$$

$$= \sum_{i=1}^{n} -\frac{1}{2}\cdot\frac{1}{\sigma}(y_i-\vec{x}_i^T\vec{\beta})^T(y_i-\vec{x}_i^T\vec{\beta}) - \frac{1}{2}\cdot\frac{1}{\tau}\vec{\beta}^T\vec{\beta} + C.$$

$$= -\frac{1}{2\sigma}(y-X\beta)^2 - \frac{1}{2\tau}\|\beta\|^2 + C$$

$$= -\frac{1}{2\sigma}\left[(y-X\beta)^2 + \frac{\sigma}{\tau}\|\beta\|^2\right] + C$$

$\max \log p(\vec{\beta}|\vec{y})$ 等价于 $\min (y-X\beta)^2 + \frac{\sigma}{\tau}\|\beta\|^2 \quad \therefore \lambda = \frac{\sigma}{\tau}$.

9. $D_l = \{(x_i,y_i)\}_{i=1}^{l}$ 已知，$D_u\{(x_i,y_i)\}_{i=l+1}^{m}$ 未知

令 $z_{ik} = \mathbb{I}(x_i, y_i=k)$，$\forall i\ l+1\le i\le m$，易知 $z_{ik} \sim B(1, p(y_i=k|x_i))$.

E-step：$E(z_{ik}) = p(y_i=k|x_i) = \dfrac{p(x_i|y_i=k)p(y_i=k)}{p(x)} = \dfrac{p(x_i|y_i=k)p(y_i=k)}{\sum_{j=1}^{n} p(x_i|y_i=j)p(y_i=j)}$

M-step：$\prod_{i=1}^{m} p(x_i,y_i) = \prod_{i=1}^{l} p(x_i,y_i)\cdot\prod_{i=l+1}^{m} p(x_i,y_i)$

$\forall i,\ l+1\le i\le m \quad p(x_i,y_i) = \prod_{j=1}^{n} p(x_i,y_i=j)^{z_{ij}}$

$$\prod_{i=1}^{m} p(x_i,y_i) = \prod_{i=1}^{l} p(x_i,y_i)\cdot\prod_{i=l+1}^{m}\prod_{j=1}^{n} p(x_i,y_i=j)^{z_{ij}}$$

$$\mu_i = \frac{\sum_{j=1}^{l} x_j \mathbb{I}(y_j=i)}{\sum_{j=1}^{l}\mathbb{I}(y_j=i)} + \frac{\sum_{j=l+1}^{m} x_j\cdot z_{ij}}{\sum_{j=l+1}^{m} z_{ij}} + \cdots \qquad \Sigma_i = \frac{1}{\sum_{j=1}^{l}\mathbb{I}(y_j=i)}\sum_{j=1}^{l}(x_j-\mu_i)^2\mathbb{I}(y_j=i) + \frac{1}{\sum_{j=l+1}^{m} z_{ij}}\sum_{j=l+1}^{m}(x_j-\mu_i)^2\cdot z_{ij}$$

$$p(y=i) = \frac{\sum_{j=1}^{l}\mathbb{I}(y_j=i)}{l} + \frac{\sum_{j=l+1}^{m} z_{ij}}{m-l}$$

10. (1)

$$p(\omega_1|x) = \frac{p(x|\omega_1)\cdot p(\omega_1)}{p(x)}$$

$$p(\omega_2|x) = \frac{p(x|\omega_2)\, p(\omega_2)}{p(x)}$$

$$p(\omega_1|x) > p(\omega_2|x)$$

$$p(x|\omega_1)\, p(\omega_1) > p(x|\omega_2)\, p(\omega_2)$$

$$\frac{p(x|\omega_1)}{p(x|\omega_2)} - \frac{p(\omega_2)}{p(\omega_1)} > 0$$

$$d(x) = \frac{p(x|\omega_1)}{p(x|\omega_2)} - \frac{p(\omega_2)}{p(\omega_1)} = 0$$

$$d(x) = \frac{0.2}{0.4} - \frac{0.9}{0.1} = -7 < 0$$

$$\therefore x \in \omega_2.$$

(2)

$$r_1 = L_{11}\cdot p(x,\omega_1) + L_{12}\cdot p(x,\omega_2)$$

$$= 0\cdot p(x|\omega_1)p(\omega_1) + 6\cdot p(x|\omega_2)p(\omega_2)$$

$$= 0 + 6\times 0.4\times 0.1$$

$$= 0.24$$

$$r_2 = L_{22}\cdot p(x,\omega_2) + L_{21}\, p(x,\omega_1)$$

$$= 0\cdot p(x|\omega_2)p(\omega_2) + 1\times p(x|\omega_1)p(\omega_1)$$

$$= 0 + 1\times 0.2\times 0.9$$

$$= 0.18$$

$$r_2 = 0.18 < 0.24 = r_1$$

$$\therefore x \in \omega_2$$

中国科学院大学

试题专用纸

课程编号：[illegible]
课程名称：模式识别与机器学习
任课教师：黄庆明、兰艳艳、郭嘉丰、山世光

注意事项：
1.考试时间为 120 分钟，考试方式 闭 卷；
2.全部答案写在答题纸上；
3.考试结束后，请将本试卷和答题纸、草稿纸一并交回。

1. （8 分）试阐述线性判别函数的基本概念，并说明既然有线性判别函数，为什么还需要非线性判别函数？假设有两类模式，每类包括 6 个 4 维不同的模式，且良好分布。如果它们是线性可分的，问权向量至少需要几个系数分量？假如要建立二次的多项式判别函数，又至少需要几个系数分量？（设模式的良好分布不因模式变化而改变）

2. （8 分）简述 SVM 算法的原理。如果使用 SVM 做二分类问题得到如下结果，分别应该采取什么措施以取得更好的结果？并说明原因。
(1) 训练集的分类准确率 90%，验证集的分类准确率 90%，测试集的分类准确率 88%；欠拟合
(2) 训练集的分类准确率 98%，验证集的分类准确率 90%，测试集的分类准确率 88%。过拟合

3. （8 分）请从两种角度解释主成分分析（PCA）的优化目标。

4. （8 分）请给出卷积神经网络 CNN 中卷积、Pooling、RELU 等基本层操作的含义。然后从提取特征的角度分析 CNN 与传统特征提取方法（例如 Gabor 小波滤波器）的异同。

5. （10 分）用线性判别函数的感知器赏罚训练算法求下列模式分类的解向量，并给出相应的判别函数。

ω_1: $\{(0\ \ 0)^T,\ (0\ \ 1)^T\}$

ω_2: $\{(1\ \ 0)^T,\ (1\ \ 1)^T\}$

6. （10 分）试述 K-L 变换的基本原理，并将如下两类样本集的特征维数降到一维，同时画出样本在该空间中的位置。

ω_1: $\{(-5\ \ -5)^T,\ (-5\ \ -4)^T,\ (-4\ \ -5)^T,\ (-5\ \ -6)^T,\ (-6\ \ -5)^T\}$

ω_2: $\{(5\ \ 5)^T,\ (5\ \ 6)^T,\ (6\ \ 5)^T,\ (5\ \ 4)^T,\ (4\ \ 5)^T\}$，

其中假设其先验概率相等，即 $P(\omega_1)=P(\omega_2)=0.5$。

（12 分）请解释 AdaBoost 的基本思想和工作原理，写出 AdaBoost 算法

8. （12 分）选择埃尔米特多项式，其前面几项的表达式为

$H_0(x)=1$，　$H_1(x)=2x$，　$H_2(x)=4x^2-2$，

$H_3(x)=8x^3-12x$，　$H_4(x)=16x^4-48x^2+12$

试用二次埃尔米特多项式的势函数算法求解以下模式的分类问题

ω_1：$\{(0\ \ 1)^T,\ (0\ \ -1)^T\}$

ω_2：$\{(1\ \ 0)^T,\ (-1\ \ 0)^T\}$

9. （12 分）已知以下关于垃圾邮件的 8 条标注数据，A、B 为邮件的 2 个特征，Y 为类别，其中 Y=1 表示该邮件为垃圾邮件，Y=0 表示该邮件为正常邮件。请依此训练一个朴素贝叶斯分类器，并预测特征为“A=0，B=1”的邮件是否为垃圾邮件。

序号	1	2	3	4	5	6	7	8
A	0	0	1	1	1	1	1	1
B	0	0	0	0	0	0	1	1

10. （12 分）假设有 3 个罐子，每个罐子里都装有红、黑两种颜色的弹珠。按照下面的方法取弹珠：开始，以概率 π 随机选取 1 个罐子，从这个罐子以概率 B 随机取出一个弹珠，记录其颜色后，放回；然后，从当前盒子以概率 A 随机转移到下一个盒子，再从这个盒子里以概率 B 随机抽出一个球，记录其颜色，放回；如此重复 3 次，得到一个弹珠的颜色观测序列：O=（红，黑，红）。请用前向传播算法计算生成该序列的概率 $P(O|\{A, B, \pi\})$。

$$\pi=[0.4,0.4,0.2]^T \qquad A=\begin{array}{c|ccc} & \text{罐子1} & \text{罐子2} & \text{罐子3} \\ \hline \text{罐子1} & 0.3 & 0.5 & 0.2 \\ \text{罐子2} & 0.2 & 0.3 & 0.5 \\ \text{罐子3} & 0.1 & 0.4 & 0.5 \end{array} \qquad B=\begin{array}{c|cc} & \text{红} & \text{黑} \\ \hline \text{罐子1} & 0.7 & 0.3 \\ \text{罐子2} & 0.5 & 0.5 \\ \text{罐子3} & 0.4 & 0.6 \end{array}$$