中国科学院大学

试题专用纸

考试日期：2017.1.11

课程编号：092M4002H

课程名称：模式识别

任课教师：刘成林

姓名__________ 学号__________ 成绩__________

1. （16 分）本题有两小题。

（1）（8 分）对一个 c 类分类问题，假设各类先验概率为 $P(\omega_i), i=1,\dots,c$，条件概率密度为 $P(\mathbf{x}|\omega_i), i=1,\dots,c$（这里 $\mathbf{x}$ 表示特征向量），将第 j 类模式判别为第 i 类的损失为 λ_{ij}。请写出贝叶斯最小风险决策和最小错误率决策的决策规则；

（2）（8 分）在 2 维特征空间，两个类别分别有 4 个样本：$X_1=\{(3,4), (3,8), (2,6), (4,6)\}$，$X_2=\{(3,0),(3,-4), (1,-2), (5,-2)\}$，假设两个类别的概率密度都为高斯分布（正态分布）$N(\mu_i, \Sigma_i)$，请写出两个类别的最大似然估计参数值$(\mu_i, \Sigma_i)$。进一步，假设两个类别先验概率相等，请写出分类决策面的公式。

2. （12 分）表示模式的特征向量 $\mathbf{x}\in R^d$，对一个 c 类分类问题，假设各类先验概率相等，每一类条件概率密度为高斯分布。

（1）（6 分）请写出在下面两种情况下的最小错误率决策判别函数：(a)类协方差矩阵不等；(b)所有类协方差矩阵相等。

（2）（6 分）当 $c=2$, $P(\omega_1)= P(\omega_2)$，两类概率密度均为高斯分布且 $\Sigma_1=\Sigma_2$，请写出贝叶斯分类决策面和贝叶斯错误率的公式。

3. （10 分）特征空间中概率密度的非参数估计近似为 $p(\mathbf{x})=\dfrac{k/n}{V}$，其中 V 为 x 周边邻域的体积，k 为邻域内样本数，n 为总样本数。基于此定义，

（1）（5 分）请说明 Parzen 窗估计和 k-近邻（k-NN）估计的区别。

（2）（5 分）对于 c 个类别，基于 k-NN 概率密度估计进行贝叶斯分类，请写出各个类别的后验概率 $p(\omega_i|\mathbf{x})$ 并证明之。

4. （10 分）现有四个来自于两个类别的二维空间中的样本，其中第一类的两个样本分别为$(3,2)^T$和$(2,2)^T$，第二类的两个样本分别为$(1,0)^T$和$(2,0)^T$。这里，上标 T 表示向量转置。若采用规范化增广样本表示形式，并假设初始的权向量 $a=(1,0,0)^T$，其中向量 a 的第三维对应于样本的齐次坐标。同时，假定梯度更新步长 η_k 固定为 1。试利用批处理感知器算法求解线性判别函数 $g(y)=a^Ty$ 的权向量 a。（注："规范化增广样本表示"是指

学　号：
姓　名：

对齐次坐标表示的样本进行规范化处理）

5. （共 10 分）

（1）（6 分）现有八个二维空间中的样本：$\mathbf{x}_1 = (-4,1)^T$、$\mathbf{x}_2 = (-2,1)^T$、$\mathbf{x}_3 = (-4,-1)^T$、$\mathbf{x}_4 = (-2,-1)^T$、$\mathbf{x}_5 = (6,1)^T$、$\mathbf{x}_6 = (4,1)^T$、$\mathbf{x}_7 = (6,-1)^T$、$\mathbf{x}_8 = (4,-1)^T$。这里，上标 T 表示向量转置。假定初始聚类中心分别为$(-2,0)^T$和$(4,0)^T$，请采用 K 均值聚类算法对上述八个样本进行聚类，写出聚类计算过程和聚类结果。

（2）（4 分）假定一组样本是从一个混合高斯密度函数中随机采样而得到的。请描述：在对混合高斯密度函数的参数进行估计的过程中，在哪些条件下可以导出 K 均值聚类算法。

6. （共 15 分）

（1）（9 分）针对多层前馈神经网络，请给出误差反向传播算法（即 BP 算法）的原理；结合三层网络给出有关权重更新的公式，并用文字描述所述公式的含义。

（2）（6 分）请描述自组织映射网络的构造原理；针对网络训练，请给出自组织算法的主要计算步骤。

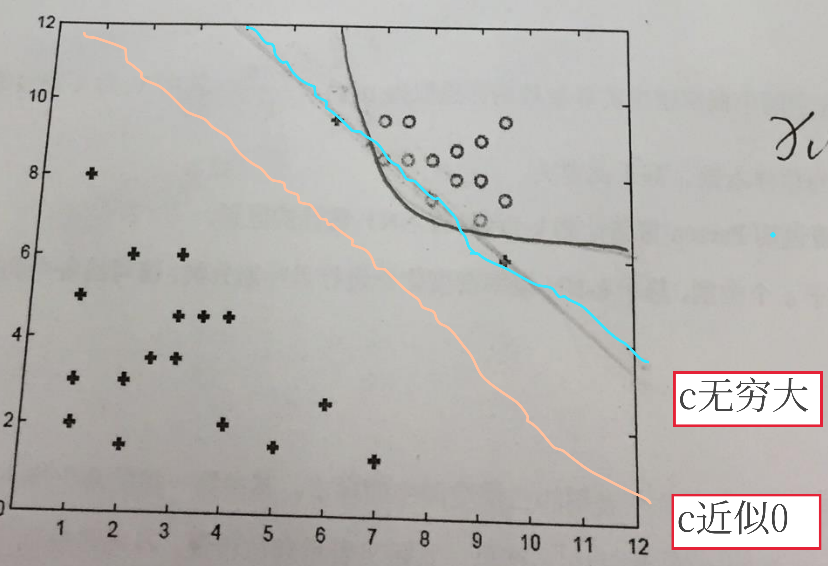
7. （共 12 分）

（1）（4 分）简述 LDA（线性判别分析）的主要思想；
（2）（4 分）基于上述思想，给出两类问题的 LDA 目标函数；
（3）（4 分）最优化上述目标函数，得到 LDA 结果。

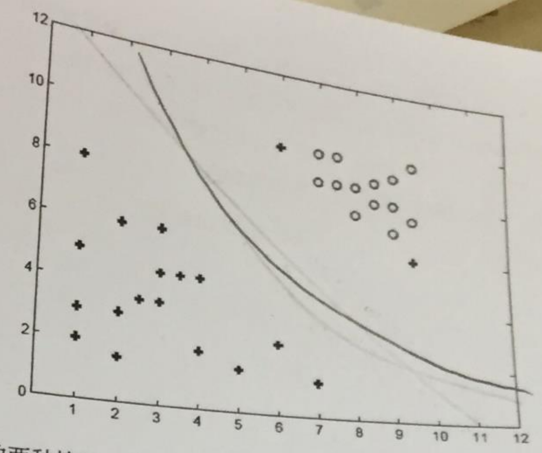
8. （15 分））解答下面关于支持向量机（SVM）的问题，每小题 3 分。

如图给定一批训练数据（**有噪声**），要训练 SVM 分类器（二分类）对未来测试数据进行分类。假设判别函数使用**二阶多项式核函数**。根据 SVM 原理，软间隔惩罚参数 C 会影响决策边界的位置。在下列各小题中，定性画出分类决策边界，**并用一两句话说明产生如此边界的理由**。

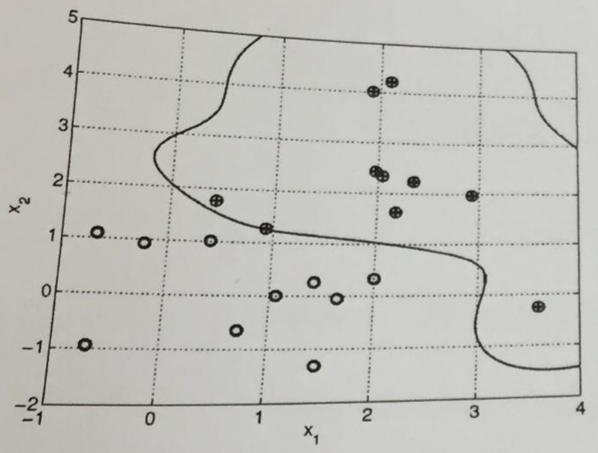
（1）当参数 C 取值特别大时（比如 C → ∞），（**在答题纸上**）画出相应的分类决策边界，并说明理由。



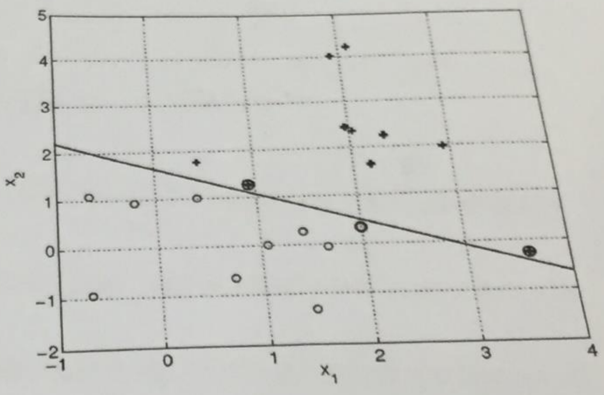
（2）当参数 C 取值特别小时（比如 C ≈ 0），（**在答题纸上**）画出相应的分类决策边界，并说明理由。



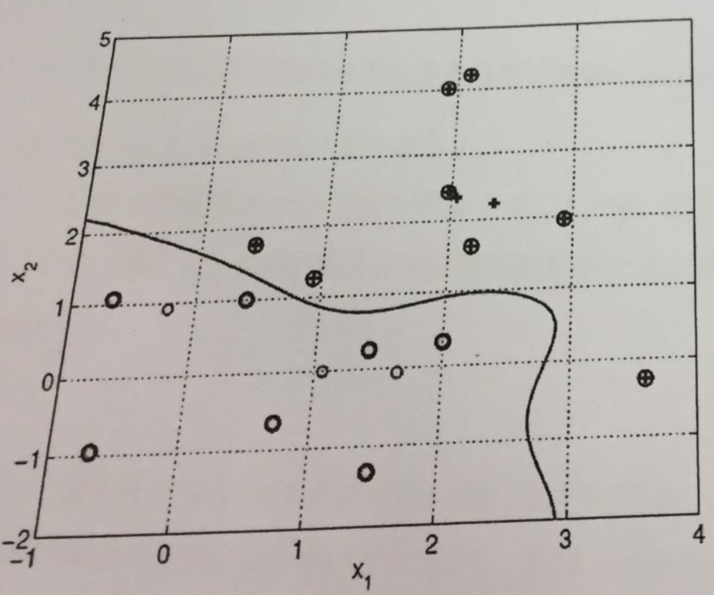
（3）对于（1）和（2）中的两种情形，你认为哪一种会在测试数据上表现出较好的性能，为什么？



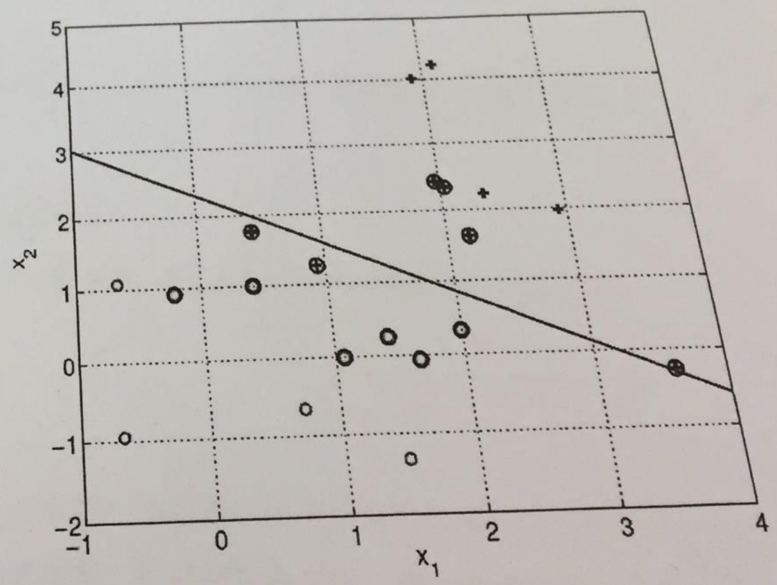
(a)



(b)



(c)



(d)

学　　号：

姓　　名：

考试科目：

上图（a-d）画出了使用相应参数和模型设置的 SVM 决策边界（二分类问题，圆圈为正样本，十字为负样本，被加粗了的圆圈和十字为支持向量 support vectors），这些设置包括不同的核函数（kernel）、不同的松弛变量（slack variable）和是否使用偏置项（bias/offset）。若干用以产生这些边界的数学模型也一并给出。请给每个数学模型找到其正确的决策边界匹配，**将选项填入相应的空白处，并说明理由。**

（4）下述模型与图___c___匹配（从“a，b，c，d”中选择），给出原因。

$$\min \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{m}\xi_i$$
$$s.t.\ \xi_i \geq 0,\ y_i(w^T x_i + b) \geq 1 - \xi_i,\ i = 1,2,\ldots,m$$
$$where\ C = 0.1$$

（5）下述模型与图___c___匹配（从“a，b，c，d”中选择），给出原因。

$$\max \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j K(x_i, x_j)$$
$$s.t.\ 0.1 \geq \alpha_i \geq 0.\ \ i = 1,2,\ldots,m,$$
$$where\ K(x_i, x_j) = \exp(-\|x_i - x_j\|^2)$$

RBF核