**《人工智能与模式识别》平时作业**

**第一章作业**：(共5分)（简答题 ×3）

1. 简述模式识别的目标和原理。(1分)
2. 简述设计一个完整的模式识别系统需要有哪些环节？(2分)
3. 超市的商品上贴有条形码，结账时收银员通过扫描条形码获取商品信息，在此过程中的条形码识别属于模式识别问题吗？请说明原因。(2分)

**第二章作业**：(共5分)

（单选（0.5分）×2，简答题 ×2）

1.【单选】 关于分类模型的复杂度与泛化能力的关系，下列描述正确的是（ ）

A. 分类模型越简单，模型的泛化能力越好

B. 分类模型越复杂，模型的泛化能力越好

C. 为提高分类模型的泛化能力，模型既不能太简单也不能太复杂

D. 简单的分类模型会产生更好的泛化能力

2. 【单选】关于模型的泛化性能的论述中，( ) 是正确的论述。

A．模型在训练数据集上的训练误差越小，表明其泛化性能越好

B. 使用更多的训练数据可以提高模型的泛化性能

C. 训练集大小不变时，使用复杂的模型一定能提高模型的泛化性能

D. 模型的泛化性能定义为模型在测试数据集上的性能

1. 简述什么是“过拟合”现象，它与模型的泛化能力有什么关系？(1分)

4. 详述设计一个车牌识别系统的整个流程。(3分)

**第三章作业**：(共10分)

（单选（0.5分）×4，多选（1分）×3，简答题（0.5分）×2，大题（4分）×1）

1.【单选】用一对多的构建方式来求解三分类问题时，在什么情况下是可以做出分类决策的？

A. 三个判别函数值都小于0 B. 仅有一个判别函数值大于0

C. 有两个判别函数值大于0 D. 三个判别函数值都大于0

2.【单选】用一对一的构建方式来求解三分类问题时，做出多分类决策的原则是？

A. 没有冲突的原则 B. 最大值的原则

C. 多数投票的原则 D. 三个判别函数值都大于0

3.【单选】梯度下降算法中，学习步长的取值情况为：

A.可以取小于0的数 B.可以取值为0

C.取值任意 D.只能取正数

4.【单选】感知器算法用什么方法求解准则函数的最优值？

A.最小均方误差 B.梯度下降法

C.最大均方误差 D.平均值法

5.【多选】下列选项中，属于感知机的特点的是？

A. 感知机将输入信号加权求和与阈值比较，当信号加权和大于阈值时输出1，当信号加权和小于阈值时输出-1

B.单个感知机可以解决非线性分类问题

C.感知机具有多路输入、单路输出

D.单个感知机只能解决二分类问题

6.【多选】下列选项中，哪些因素会影响到感知器算法中随机梯度下降法的求解结果？

A.学习规则 B.学习步长

C.初始权向量设置 D.样本处理顺序不同

7.【多选】通过构建多个线性判别函数可以解决多分类问题，以下哪种表述是正确的？

1. 一对多的构建方式存在不可识别区域较多的问题

B.一对一的构建方式中不可识别的区域比一对多方式的更多

C.一对一的构建方式需要训练的线性判别函数的数目较多

D.最大值组合方式不存在不可识别区域，其构成的多分类器的整体性能比较好

1. 简述线性分类器训练的一般思路是什么？
2. 简述广义线性化的思想。

10.给定一个训练集，其中正例样本为，，负例样本为，请用感知机算法求出能正确分类该数据集的分类模型：

，其中， .

请使用随机梯度下降法更新，初始化为零向量，取学习步长η=1，写出详细的计算过程。

**第四章作业**：(共12分)

（单选（0.5分）×4，多选（1分）×3，大题（7分）×1）

1.【单选】硬间隔最大化思想中的“间隔”指的是什么？

A. 训练数据集中所有样本点到决策超平面的垂直距离的平均距离

B. 训练数据集中所有样本点到决策超平面的垂直距离中的最大距离

C. 样本点到决策超平面的垂直距离

D. 训练数据集中所有样本点到决策超平面的垂直距离中的最小距离

2.【单选】关于“间隔最大化”的动机，以下最准备的描述是哪一个？

A. 找到一个超平面，以足够大的置信度将所有训练样本正确分开

B. 找到一个将训练样本正确分开的超平面

C. 将训练样本正确分开

D. 无

3.【单选】线性支持向量机对误分类的容忍程度与其惩罚系数有什么关系？

A. 二者没有联系

B. 二者之间存在联系，但是不能找到联系间存在的规律

C. 惩罚系数越大，容忍度越低

D. 惩罚系数越大，容忍度越高

4.【单选】在使用非线性SVM分类算法前，需要提前确定（ ）

A. 将输入数据映射到多少维的高维空间中

B. 非线性映射

C. 数据映射到高维空间中后是否是线性可分的

D. 核函数和惩罚参数

5.【多选】原始问题和对偶问题的最优解一定满足KKT条件，请问KKT条件包括哪些条件？( )

A. 拉格朗日函数关于原问题变量的梯度在最优解处为零向量

B. 对偶问题的约束

C. 对偶互补条件

D. 原始问题的等式和不等式约束

6.【多选】关于线性可分的SVM中的支持向量，下列描述正确的是

A. 将除支持向量以外的样本点删除，也不影响分类决策边界

B. 支持向量离决策超平面的距离最近

C. 支持向量的数目一般比较少

D. 在决定分类决策超平面的位置时，只有支持向量起作用，其它样本点不起作用

7.【多选】关于软间隔SVM，以下描述正确的有：

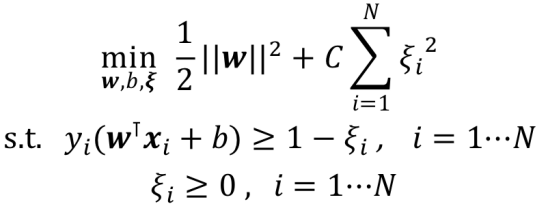
A. 能解决由于少量的噪声或异常点引起的线性不可分问题

B. 不强迫所有的训练样本都被正确分类，允许一些样本被错误分类

C. 能解决真正的非线性分类问题

D. 比硬间隔最大化的SVM适用范围更广

8. 线性支持向量机的优化问题还可以表示为如下形式：



1. 请推导出该优化问题的对偶问题。
2. 假设对偶问题最优解为，请利用KKT条件构造出原始问题的最优解。

**第五章作业：**（共20分）

（单选（0.5分）×6，多选（1分）×3，大题（14分））

1.【单选】最大似然估计中，使用对数似然函数是为了\_\_\_\_。

A.提高似然函数灵敏度 B.将求极大值问题转化为求极小值问题

C.便于对似然函数求和 D.便于对似然函数求导

2.【单选】已知要估计的概率密度函数是多峰的，则选择\_\_\_\_法能够得到一个相对较好的密度曲线估计。

A.贝叶斯估计 B.使用正态窗函数的parzen窗

C. Kn近邻法 D.最大似然估计

3.【单选】下面关于贝叶斯分类器的论述中，（ ）是错误的。

A. 贝叶斯分类器的分类决策边界可以是线性的，也可以是非线性的

B. 贝叶斯分类器是最小化分类错误率的分类器

C. 最小化错误率的贝叶斯分类器等价于最小化风险的贝叶斯分类器

D. 最小化错误率的贝叶斯分类器没有考虑不同的错分类的代价

4.【单选】以下关于K近邻分类器的描述错误的是\_\_\_\_\_.

A. K近邻分类器是一个非线性分类器

B. K近邻分类器可以直接用于多类的分类问题。

C. SVM和贝叶斯分类器需要提前训练好分类器才能用于分类，而K近邻分类器不需要提前训练分类器。

D. K近邻分类器的分类结果不受距离函数的选择的影响。

5.【单选】在2维的坐标系中，距离原点的ℓ\_1范数距离(曼哈顿距离)等于1的点的轨迹是\_\_\_\_。

A. 以原点为中心，以两个坐标轴为对角线，边长为根号2的正方形

B. 以原点为中心，边长为2，各个边与坐标轴平行的正方形

C. 以原点为中心，半径为1的圆

D. 以原点为中心，边长为根号2，各个边与坐标轴平行的正方形

6.【单选】以下关于特征归一化/标准化的描述错误的是\_\_\_\_.

A.特征归一化的目的是近似地均衡每维特征的影响，使每维的特征在距离计算中有近似相同的作用。

B.特征归一化在任何的数据集上都是适用的。

C.对某一维度的特征进行z-score标准化后，该维度特征的均值为0，方差为1。

D.对数据集进行min-max标准化会将每个维度上的特征值变换到相同的取值范围内。

7.【多选】关于距离，下列说法正确的是\_\_\_\_\_.

A. 欧式距离就是两点之间的直线距离。

B. 曼哈顿距离是两个点在所有维度上的距离之和，其中在每个维度上的距离为两个点在该维度上的坐标值之差的绝对值。

C. ℓ\_∞范数距离又称为城市街区距离。

D. ℓ\_∞范数距离是两个点在各个维度上距离的最大值。

8.【多选】贝叶斯分类器的训练，就是利用训练样本集估计出\_\_\_\_。

A.全概率 B.类条件概率/概率密度

C.后验概率 D.先验概率

9.【多选】下列描述正确的是\_\_\_\_。

A.非参数化的估计方法直接对概率密度函数p(x|ω\_j)本身进行估计，而不必假设p(x|ω\_j)具有某个确定的分布形式。

B.贝叶斯估计法把待估计的参数θ看作是随机量，服从某个已知的先验分布。

C.参数化方法假定p(x|ω\_j)具有某种确定的分布形式，但是分布的参数未知，然后利用训练集对分布的参数进行估计。

D.利用最大似然估计法对某类的类条件概率密度函数p(x|ω\_j)进行估计时，需要假设该类中的样本是独立同分布的。

10. 已知一个样本集是一维的，包含和两个类别的样本，每类样本出现的先验概率相同：，且类条件概率密度函数为：

(1)根据贝叶斯分类规则，判断应该将新样本预测为哪一类？（2分）

(2)记表示一个样本的真实类别为时，将其预测为类所产生的损失，且，，，根据最小化风险贝叶斯分类规则，判断将新样本预测为哪一类？（2分）

11. 在两类三维空间的分类问题中，第一类的特征向量样本服从三维正态分布，第二类特征向量服从三维正态分布，其中，，，且两类特征向量出现的概率相等。

(1)试写出将贝叶斯分类规则通过判别函数表示法表示时，相应的两个判别函数的方程和分类决策面的方程各是什么。（2分）

(2)根据贝叶斯分类规则判断应该将一个样本分为哪一类？（2分）

12. 假设一个一维样本集D={,}是根据参数θ未知的伯努利分布Bernoulli(θ)采样得到的，即， 其中*x*=0或1，0≤θ≤1，请采用最大似然估计法给出θ的估计，写出详细的推导步骤。（3分）

13. 已知二维空间中的3个点，，，试求在分别取1，2和时，在范数距离下的最近邻点。（3分）

**第六章作业：**（共10分）（多选（1分）×1，大题（9分））

1.【多选】下列描述正确的有\_\_\_\_\_.

A. PCA寻找的是能够有效分类的方向，而FDA寻找的是能够有效表示数据的方向。

B. FDA寻找到的最优投影方向满足：不同类的投影点的均值之间的差距最大，同类的投影点的分散程度(方差)最小。

C. 数据集的本征维度是表示数据集所需要的最小变量数，与数据集的原始维度不同，一般地，本征维度大于原始维度。

D. PCA是无监督的降维方法，不考虑样本集的类标信息，而FDA则是有监督的降维方法，需要知道样本集的类标信息。

2. PCA算法的主成分e\_1, e\_2,..., e\_k之间有什么关系? 每个主成分的模是多少? (1分)

3. 给定d维的样本集D={x\_1, x\_2,⋯, x\_n}，μ为样本集的均值向量，已知样本的**协方差矩阵**定义为, 这里符号表示矩阵的转置运算。在PCA中，**散布矩阵**定义为，可以看到散布矩阵是协方差矩阵的倍。试分析在PCA用协方差矩阵来代替散布矩阵会有什么样的影响，会影响PCA降维的结果吗？(2分)

4. 设原始数据集D={x\_1, x\_2,⋯, x\_n}是d维的，采用PCA降维后，变成k维的，其中kd。设PCA找到的k个主成分依次为e\_1, e\_2,..., e\_k，则任意的d维样本x进行PCA降维后，在第r个主成分上的坐标为：，其中。因此，第i个样本在第r个主成分上的投影点可以表示为，所以，所有样本在第r个主成分上的投影点依次为y\_1, y\_2, ...y\_n，共n个投影点。

①计算这n个投影点的均值; (1分)

②计算这n个投影点的方差；（注意使用样本方差公式）（4分）

③利用上面的结果来说明，PCA的第一主成分就是所有样本投影到该方向上的投影点的方差最大的方向。（1分）

1. **作业：**（共12分）（多选（1分）×2，大题（10分））

1.【多选】下列属于无监督学习方法的是\_\_\_\_\_ 。

A. PCA B. FDA C. k-means D. kNN。

2.【多选】下面关于聚类算法的描述，\_\_\_\_\_ 是错误的。

A. k-medoids聚类的结果易受噪音或离群点的影响。

B. k-means算法比k-medoids算法更高效。

C. k-means只能发现球状的簇，而DBSCAN能够发现任意形状的簇。

D.算法DBSCAN对噪声/离群点比较敏感。

3. 简述有监督学习和无监督学习的区别。(2分)

4. 给定一个样本集{(0,0), (0,1), (1,0), (2, 1), (2, 3), (3, 4)}, 试用基于欧式距离的k-均值聚类算法将该数据集聚成2类，要求初始的聚类中心选择c1=(0,0) , c2 = (0,1)。请写出详细的计算过程，给出算法每次迭代结束后，将哪些样本点聚在一个簇中，并注明每次迭代的聚类中心。（8分）

**第八章作业：**（共10分）

1. 对某地区的人种情况进行调查，得到如下的一张表：



其中表中每一行信息表示一个人员的信息，包括眼睛颜色和头发颜色，最后一列表明该人员属于哪种人种。根据该表构建一棵决策树，能根据一个人的眼睛和头发的颜色来判断其所属人种。要求在选择划分属性时，使用信息增益的准则，计算熵时以2为底。请写出详细的构建过程，并画出最终得到的决策树。

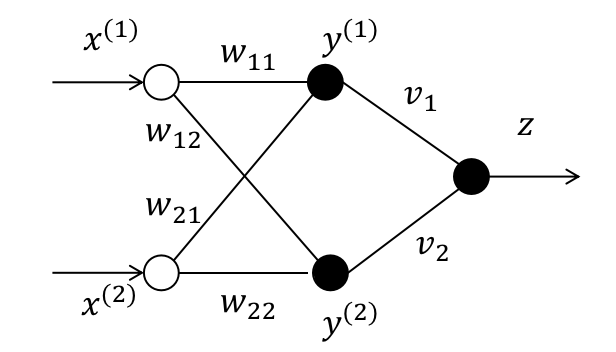
（7分）

2. 设X是一个离散型随机变量，其概率分布为：

求X的熵的最大值，其中熵以纳特为单位，求解过程按步骤计分。（3分）

1. 作业：（共8分）

1. 给定如下图所示的前馈神经网络：

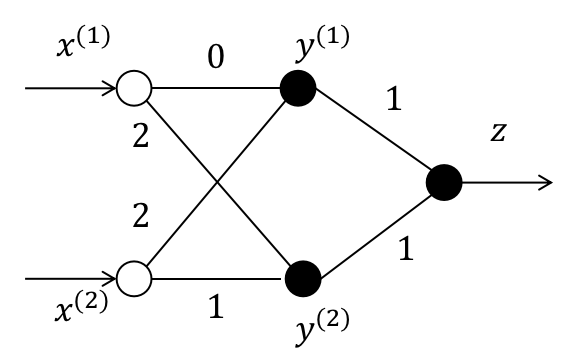


输入为和，隐层输出为和，输出层的输出为，输入层到隐层的连接权和隐层到输出层的连接权见图所示，每个功能神经元的阈值均设置为0，且每个功能神经元采用RELU激活函数：

① 根据信息前向传播机制，用图中符号给出，和的表达式(1分)

② 根据误差反向传播机制，推导出连接权，，，，，的更新公式 (5分)

③ 假设上述神经网络在第k-1次学习中得到的连接权值如下图所示：



在第k次学习中，新输入的样本为(1, 0)， 其类标为1，即期望输出为1，请应用①②得出的公式，计算第k次学习后，各个连接权的值为多少？更新时请采用学习步长。(2分)

1. 作业：（共8分，选择（1分）×2，大题（6分））

1.【单选】关于集成分类器，下列描述错误的是（ ）。

A. Adaboost算法训练的基分类器之间存在强依赖关系，而Bagging算法训练的基分类器之间不存在依赖关系

B. Adaboost算法中，基分类器只能以串型的方式训练，而Bagging算法中，基分类器的训练可以并行执行

C. Adaboost算法通过改变训练样本的权值分布来构建不同的训练数据集，继而训练各个基分类器，而Bagging算法是通过bootstrap自助抽样法来得到不同的训练数据集

D. Adaboost算法通过基分类器的多数投票来得到最终分类结果，而Bagging算法是通过各个基分类器的加权组合得到最终分类结果

1. 【多选】下面哪些分类器属于不稳定的分类器（ ）。
2. 决策树 B. 支持向量机 C.神经网络 D. k近邻分类器

3. 简述BP神经网络如何接受带权值的训练样本集来训练神经网络的参数。(1分)

4. 一个已经训练好的BP神经网络模型对待识别的实例***x***的输出为向量，则网络应该将***x***预测为哪一类？ (1分)

5. 设Adaboost算法总共训练了M个弱学习器，且以BP神经网络作为弱学习算法。在预测一个实例***x***的类标时，若第个训练好的弱学习器对***x***的预测为向量，且该弱学习器的决策权重为，则Adaboost算法应该将***x***预测为哪一类？请用数学公式表示。 (2分)

6. 当Bagging算法以BP神经网络作为弱学习算法时，如果第个训练好的弱学习器将***x***预测为类，则Bagging算法应该将***x***预测为哪一类？ (2分)