北京航空航天大学研究生课程试卷 **A**

2022－2023 学年第一学期期末试卷

学号_____ 姓名_____ 成绩_____

考试日期： 2022 年 12 月 22 日，下午 15:50－17:50

考试科目：《 机器学习 》（A 卷）

注意事项：1、请大家仔细审题

2、不能违反考场纪律

一.Minsky与Papert指出：感知机因为是线性模型，所以不能表示复杂的函数，如异或。异或函数输入$X_1=(1,-1)^T$，$X_2=(-1,1)^T$时，输出1；输入$X_3=(1,1)^T$，$X_4=(-1,-1)^T$时，输出-1。试着使用感知机算法求出分类决策函数，并写出迭代步骤，验证感知机为什么不能表示异或。感知机模型为$f(x)=\text{sign}(b+a^Tx)$，初始解向量$a_1=(0,0)^T$，$b_1=0$，梯度下降法的步长为1（注：请按照实例下标的顺序对实例进行迭代）。（10分）

二.以下西瓜数据集包含 10 个训练样例，7 个测试样例，用来学习一棵能够预测没剖开的西瓜是不是好瓜的决策树。请根据下表，使用 ID3 算法建立决策树，并采用预剪枝策略进行剪枝。要求写出每个计算步骤，画出决策树的各层，指出叶子结点所代表的类别。注意：结点的类别标记为训练样本样例数最多的类别，如果样例数相等，则约定该结点类别标记为好瓜。（10 分）

表 1 训练样例

编号	色泽	根蒂	敲声	好瓜
1	青绿	蜷缩	浊响	是
2	乌黑	蜷缩	沉闷	是
3	乌黑	蜷缩	浊响	是
6	青绿	稍蜷	浊响	是
7	乌黑	稍蜷	浊响	是
10	青绿	硬挺	清脆	否
14	浅白	稍蜷	沉闷	否
15	乌黑	稍蜷	浊响	否
16	浅白	蜷缩	浊响	否
17	青绿	蜷缩	沉闷	否

表 2 测试样例

编号	色泽	根蒂	敲声	好瓜
4	青绿	蜷缩	沉闷	是
5	浅白	蜷缩	浊响	是
8	乌黑	稍蜷	浊响	是
9	乌黑	稍蜷	沉闷	否
11	浅白	硬挺	清脆	否
12	浅白	蜷缩	浊响	否
13	青绿	稍蜷	浊响	否

三.说明K均值算法和混合高斯模型的异同点。简述如何使用EM算法估计混合高斯模型的参数。（10分）

四.（1）简述支持向量机的基本思想，并写出模型表达式。说明支持向量机如何处理非线性可分的数据。（5 分）

（2）简述直推式支持向量机的基本思想，并写出模型表达式。说明直推式支持向量机如何避免穷举未标记样本的各种标记指派。（5分）

五.在线性分类器设计的最小二乘准则中，b的选取具有多样性。试证明在二分类问题中，当$b=[\underbrace{N/N_1,N/N_1,N/N_1,...}_{N_1\text{项}},\underbrace{...N/N_2,N/N_2,N/N_2}_{N_2\text{项}}]^T$（$N$为样本总数，$N_1$为第一类样本个数，$N_2$为第二类样本个数），此时基于最小二乘准则求解出的线性分类面等价于基于Fisher准则求解出的线性分类面。（10分）

六.存在一组二维样本：

$$X_1 = [6,3]^T, X_2 = [0,3]^T, X_3 = [-3,-6]^T, X_4 = [0,0]^T, X_5 = [-3,0]^T$$

（1）使用主成分分析（PCA）的方法，将这五个点降维到一维空间。（6 分）

（2）简述使用 PCA 和 LDA（Fisher 准则）的基本思想和这两个方法的异同。（4 分）

七.（1）分别以 Adaboost 和 bagging 算法为例解释串/并行集成方法原理。（6 分）

（2）请简述随机森林算法中如何引入样本扰动和属性扰动来实现基学习器的多样性。（4 分）

八.（1）包含一层隐藏层的前馈神经网络如图 1 所示，给定训练集$D = \{(x_1, t_1), (x_2, t_2), \cdots, (x_N, t_N)\}, x_i \in \mathbb{R}^D, t_i \in \mathbb{R}^K$。隐藏层包含$M$个神经元，激活函数为$\mathrm{h}(x)$，输出层激活函数为$\sigma(x)$。$y_n$表示第$n$个样本$x_n$对应的神经网络输出向量，$y_n = [y_{n1}, \cdots, y_{nk}, \cdots, y_{nK}]^T$， 准则函数为$E_n(\mathrm{w}) = \frac{1}{2}\sum_{k=1}^{K}\{y_{nk} - t_{nk}\}^2$。试推导反向传播算法中对每一层权值参数（$\omega_{md}^{(1)}$ 与 $\omega_{km}^{(2)}$）的更新。（10 分）

（2）在训练深度神经网络时，经常采用正则化的方法来避免网络出现过拟合，请列举 4 种正则化的方法并简要说明其工作原理。（8 分）

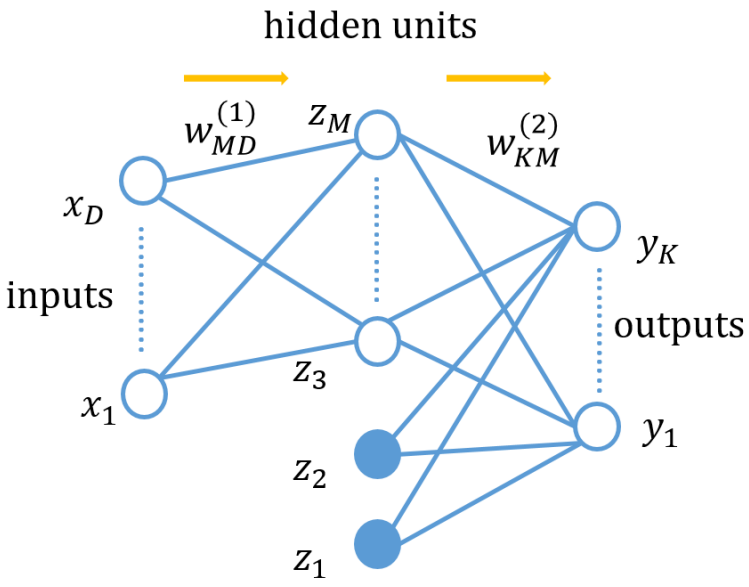

图 1

九. 请结合机器学习算法设计一套火车站人脸识别系统，必要时可以画出流程图来进行辅助说明。（12分）