

# 湖南科技大学考试试题纸（A 卷）

（2020 - 2021 学年度第二学期）

课程名称: 机器学习概论 开课单位: 数学学院 命题教师: \_\_\_\_\_

授课对象: 数学与计算科学 学院 2018 年级 信息与计算科学 1-4 班

考试时量: 100 分钟 考核方式: 考查 考试方式: 开卷

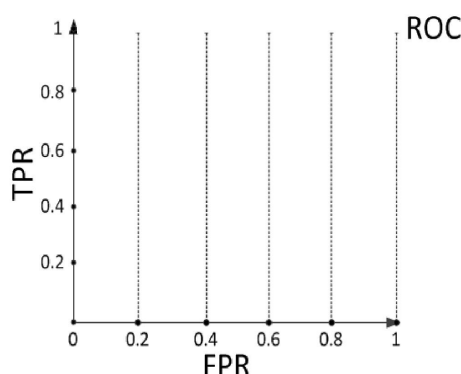
审核人: \_\_\_\_\_ 审核时间: \_\_\_\_\_ 年 \_\_\_\_\_ 月 \_\_\_\_\_ 日

## 一、计算题（本大题共 58 分，第 1 题 12 分，第 2 题 12 分，第 3 题 10 分，第 4 题 24 分）

1. 现在假设有一个训练好的二分类器对 10 个正负样本（正例 5 个，负例 5 个）预测，得分按高到低排序得到的最好预测结果为 [1, 1, 1, 0, 1, 0, 1, 0, 0, 0]

①试描绘出该预测结果的 ROC 曲线；（6 分，在答题纸上作图，在试题纸上作图不给分）

②计算 AUC（6 分）



2. 为了测定刀具的磨损速度，我们做这样的实验：经过一定时间(如每隔一小时)，测量一次刀具的厚度,得到一组试验数据如下：

顺序编号	0	1	2	3	4	5	6	7
时间 $t_i$ (小时)	0	1	2	3	4	5	6	7
刀具厚度 $y_i$ (毫米)	27.0	26.8	26.5	26.3	26.1	25.7	25.3	24.3

利用最小二乘法，根据上面的试验数据建立起  $y=at+b$  的线性回归模型。（12 分）

3. 利用三元 ECOC 码计算测试示例与 C1、C2、C3、C4 的海明距离，并根据海明距离预测出该测试示例的分类。（10 分）

	f1	f2	f3	f4	f5	f6	f7
C1	+1	-1	-1	+1	+1	-1	+1
C2	0	0	-1	0	-1	+1	0
C3	-1	+1	+1	-1	+1	-1	-1
C4	0	+1	-1	+1	0	-1	+1

三元 ECOC 码

测试示例	+1	+1	-1	+1	-1	+1	-1
------	----	----	----	----	----	----	----

4. 根据样本集 D(P86 表 4.4) 上的属性“纹理”数据，
- ①. 写出该属性上无缺失值的样例子集  $\tilde{D}$ ；(4 分)
  - ②. 计算该样例子集  $\tilde{D}$  的信息熵（保留到小数点后三位）；(8 分)
  - ③. 令  $\tilde{D}^1$ 、 $\tilde{D}^2$  与  $\tilde{D}^3$  分别表示在属性“纹理”上取值为“清晰”、“稍糊”以及“模糊”的样本子集，分别计算该三个样本子集的信息熵（保留到小数点后三位）；(9 分)
  - ④. 样本子集  $\tilde{D}$  上属性“纹理”的信息增益（保留到小数点后三位）。(3 分)。

**二、证明题（本大题共 24 分，第小题 12 分）**

1. 证明：对于图 5.7 所示的 BP 神经网络，假设隐层和输出层神经元都使用 Sigmoid 函数，有
2. 试证明样本空间中任意点  $x$  到超平面  $(w,b)$  的距离公式（6.2）

**三、论述题（18 分）**

1. 距离是泛函分析中最基本的概念之一，在《机器学习》第三章纠错输出码也应用到了海明距离和欧式距离的概念，试用数学语言阐述距离的定义。