

HOMEWORK #4（10分）

一、决策树：（4分）

下表给出了两组人的数据，每组数据分别包含4个和5个样本。

1) 用决策树方法建立该数据集的二分类器，使它能正确区分这两组人，写出建立过程。

2) 用所建分类器说明给定样例（矮，黑色，成年）是属于第几组。

组次	id	身材	发色	年龄
第一组	1	矮	金色	老人
	2	高	红色	老人
	3	高	金色	老人
	4	矮	金色	成年
第二组	1	高	黑色	儿童
	2	矮	黑色	老人
	3	高	黑色	老人
	4	高	黑色	成年
	5	矮	金色	儿童

二、朴素贝叶斯：（4分）

对于上表中的数据：

1) 用朴素贝叶斯建立二分类器，写出建立过程。

2) 用所建分类器对给定样例（矮，黑色，成年）分类。

三、K-means聚类：（2分）

给定下列 13 个数据点：

(1,3); (1,2); (2,1); (2,2); (2,3); (3,2); (5,3); (4,3); (4,5); (5,4); (5,5); (6,4); (6,5)

使用 K-means 算法对它们进行聚类。令 $k=2$ ，初始中心点为(0,4)和(6,5)，写出聚类过程。

有问题请邮件联系助教 shijx12@163.com。