**第四次作业：**

**最晚：2018.10.31提交**

**邮箱：zishuangzhao@gmail.com**

1. **选择题：**
2. 在其他条件不变的前提下，以下哪种做法容易引起机器学习中的过拟合：（）
3. 增加训练集量
4. 减少神经网络隐藏层节点数
5. 删除稀疏的特征
6. SVM 算法中使用高斯核/RBF 核代替线性核

2. 以下哪些方法不可以直接来对文本分类？（）

A．Kmeans

B.决策树

1. 支持向量机

D．KNN

3.以下描述错误的是：( )

A.SVM是这样一个分类器，他寻找具有最小边缘的超平面，因此它也经常被称为最小边缘分类器（minimal margin classifier）

B.在聚类分析当中，簇内的相似性越大，簇间的差别越大，聚类的效果就越差。

C.在决策树中，随着树中结点数变得太大，即使模型的训练误差还在继续减低，但是检验误差开始增大，这是出现了模型拟合不足的问题。

D.聚类分析可以看作是一种非监督的分类。

4一般，k-NN最近邻方法在( )的情况下效果较好

A.样本较多但典型性不好

B.样本较少但典型性好

C.样本呈团状分布

D.样本呈链状分布

5 下面哪些是基于核的机器学习算法?( )

A.Expectation Maximization

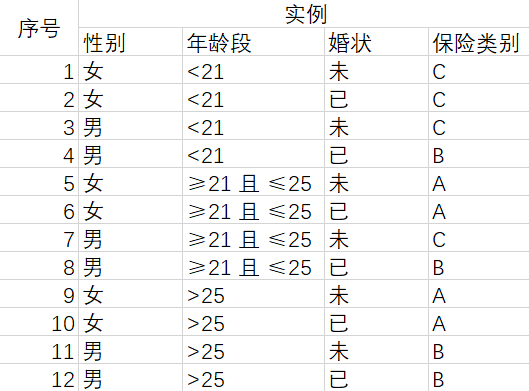
B.Radial Basis Function

C.Linear Discrimimate Analysis

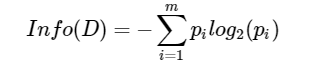
D.Support Vector Machine

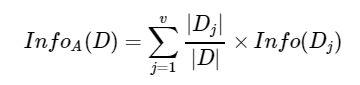
1. **简答题**
2. 参数方法的优缺点：
3. 决策树的优缺点：

**三、计算题**









**根据如上数据构建一颗决策树**

**四、编程题**

运行附录代码，查看运行结果 总结规律

提示：

从K值不同结果

每次起始点不同聚类结果不同

不同图片聚类结果对比