1.下列不属于分类方法的是（单选题）

决策树

K-means√

贝叶斯分类

神经网络

2.下列属于聚类方法的是（多选题）

A. 基于划分的聚类方法

B. 基于层次的聚类方法

C. 基于密度的聚类方法

D. 基于网格的聚类方法

E. 基于模型的聚类方法

E√

A√

C√

D√

B√

3.下列正确的KDD过程的是（单选题）

数据抽取、问题定义、数据预处理、数据挖掘、模型评估

问题定义、数据抽取、数据预处理、模型评估、数据挖掘

问题定义、数据抽取、数据预处理、数据挖掘、模型评估√

问题定义、数据预处理、数据抽取、数据挖掘、模型评估

4.下列对Apriori算法说法错误的是（单选题）

两次次扫描事务数据库√

可能产生庞大的侯选集

使用支持度来作为我们判断频繁项集的标准

采用了迭代的方法

5.下列说法错误的是（单选题）

pyspark.mllib.fpm.FPGrowthModel。这个类可以把我们之前保存的FP Tr

FP Tree算法对应的类是pyspark.mllib.fpm.FPGrowth(以下简称FPGro

Spark MLlib中，实现了两种关联算法，即我们的FP Tree和PrefixSpan

对于FPGrowth类，使用它的训练函数train主要需要输入两个参数：数据项集data，支持度阈值√

6.下列说法错误的是（单选题）

线性回归可能产生过拟合，我们需要加入正则化项

坐标轴下降法常用于Ridge回归的损失函数优化√

Lasso回归有时也叫做线性回归的L1正则化

L2正则化项，就是Ridge回归