K-近邻算法

优点：精度高、对异常值不敏感、无数据输入假定;

缺点：计算复杂度高、空间复杂度高。

适用数据范围：数值型和标称型。

k-近邻算法的一般流程

(1)收集数据：可以使用任何方法。

(2)准备数据：距离计算所需要的数值，最好是结构化的数据格式。

(3)分析数据：可以使用任何方法。

(4)训练算法：此步驟不适用于1 近邻算法。

(5)测试算法：计算错误率。

(6)使用算法：首先需要输入样本数据和结构化的输出结果，然后运行女-近邻算法判定输入数据分别属于哪个分类，最后应用对计算出的分类执行后续的处理。

K-近邻算法是分类数据最简单最有效的算法。K近邻算法是基于实例的学习，使用算法时我们必须有接近实际数据的训练样本数据。K-近邻算法必须保存全部数据集，如果训连数据集的很大，必须使用大量的存储空间。此外,由于必须对数据集中的每个数据计算距离值，实际使用时可能非常耗时。

K近邻算法的另一个缺陷是它无法给出任何数据的基础结构信息，因此我们也无法知晓平均实例样本和典型实例样本具有什么特征

决策树

优点：计算复杂度不高，输出结果易于理解，对中间值的缺失不敏感，可以处理不相关特征数据。

缺点：可能会产生过度匹配问题。

适用数据范围：数值型和标称型。

K-近邻算法可以完成很多分类任务，但是它最大的缺点就是无法给出数据的内

在含义，决策树的主要优势就就在于数据形式非常容易理解。

决策树的一般流程

(1)收集数据：可以使用任何方法。

(2)准备数据：树构造算法只适用于标称型数据，因此数值型数据必须离散化。

(3)分析数据：可以使用任何方法，构造树完成之后，我们应该检查图形是否符合预期。

(4)训练算法：构造树的数据结构

(5)测试算法：使用经验树计算错误率。

(6)使用算法：此步骤可以适用于任何监督