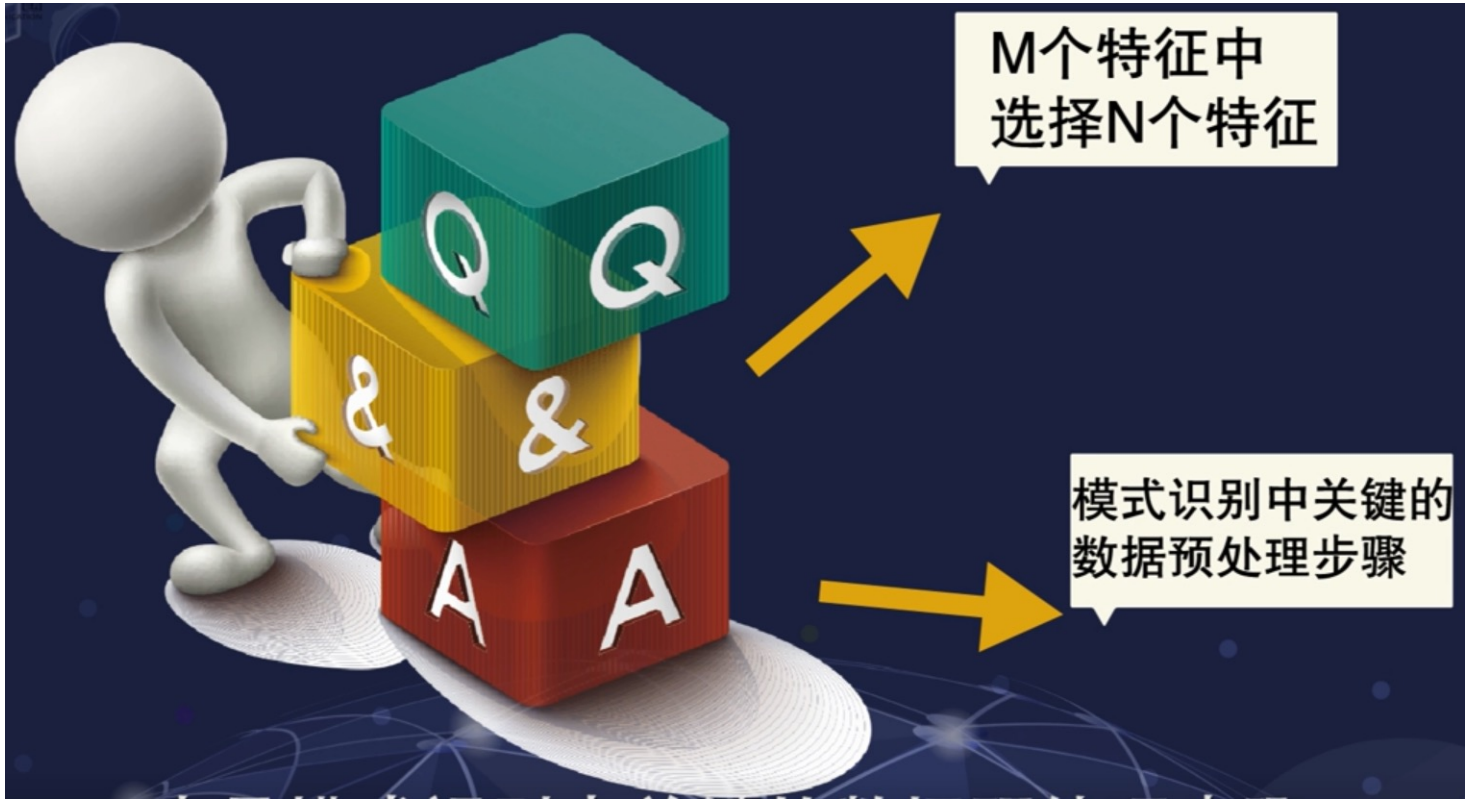
期末大作业：特征选择问题



问题描述：

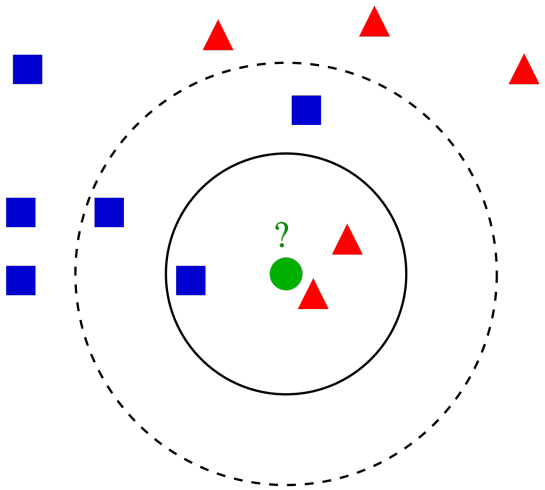
- 在机器学习和统计学中，特征选择也被称为变量选择、属性选择或变量子集选择，是指为了构建模型而选择相关特征（即属性、指标）子集的过程。
- 使用特征选择技术的关键假设是：训练数据包含许多冗余或无关的特征，因而移除这些特征并不会导致丢失信息。比如：如果一个特征本身有用，但如果这个特征与另一个有用特征强相关，且那个特征也出现在数据中，那么这个特征可能是冗余的。
- 特征选择算法可以被视为搜索技术和评价指标的结合。

算法设计要求：

- 利用2种不同的智能优化算法分别设计特征选择器，结合KNN方法解决对附件中所给出数据集wine_3cls的特征选择问题；
- 数据处理：将整个数据集分为训练样本集和测试样本集，占比分别为40%和60%左右；
- 评估指标：综合考虑选取特征数量、训练样本的拟合程度以及测试样本的泛化能力；
- 仿真实验：
  - 1）对比分析不同智能优化算法的实验结果；
  - 2）对KNN方法中K的取值进行灵敏性分析。

KNN分类算法：

- 在特征空间中，如果一个样本附近的K个最近（特征空间中最邻近）样本的大多数属于某一个类别，则该样本也属于这个类别。
- 所谓K近邻算法，即是给定一个训练数据集，对新的输入实例，在训练数据集中找到与该实例最邻近的K个实例（也就是上面所说的K个邻居），这K个实例的多数属于某个类，就把该输入实例分类到这个类中。



对于绿色圆样本点来说，

若K=3，则该样本点属于红色三角分类；

若K=5，则该样本点属于蓝色正方形分类。

验收要求：

- 验收时间：6月16日晚 6点开始分组验收；
- 验收内容：
  - 1）代码演示；
  - 2）PPT汇报5分钟左右；
  - 3）回答提问。