2018年1月4号

1. 什么是数据挖掘（KDD）（knowledge discovery database）

（https://www.zhihu.com/question/19637218）？

1. 数据挖掘主要是为了在大型数据库中发现有用的信息，在依据这些信息之间的关系进行分析，得出结果。
2. 数据处理过程：输入数据，预处理（特征选择，维归约规范化，选择数据子集），分析与挖掘，再处理（模式过滤，可视化，模式表示）。
3. 数据挖掘，不是简单的把数据呈现出来，而是要挖掘出数据之间隐藏着的不知道的关系、信息。  
   可以用，数据会说话来描述。  
   技术分类：预言：用历史预测未来；描述：了解数据中潜在的规律。
4. 相关技术包括：关联，序列发现，分类，聚集，异常检测，汇总，回归，时间序列分析。  
   可以一个个学习。
5. 数据挖掘的核心是算法

1、分类问题　　用户流失率、促销活动响应、评估用户度都属于数据挖掘的分类问题，我们需要掌握分类的特点，知道什么是有监督学习，掌握常见的分类方法:决策树、贝叶斯、KNN、支持向量机、神经网络和逻辑回归等。

　2、聚类问题　　细分市场、细分客户群体都属于数据挖掘的聚类问题，我们要掌握聚类特点，知道无监督学习，了解常见的聚类算法，例如划分聚类、层次聚类、密度聚类、网格聚类、基于模型聚类等。

　　3、关联问题　　交叉销售问题等属于关联问题，关联分析也叫购物篮分析，我们要掌握常见的关联分析算法：Aprior算法、Carma算法，序列算法等。

　4、预测问题　　我们要掌握简单线性回归分析、多重线性回归分析、时间序列

<https://www.zhihu.com/question/20751219>

**如何系统地学习数据挖掘？**

。