L3 Thinking

1、关联规则中的支持度、置信度和提升度代表的什么，如何计算

答：支持度代表的是物品（组合）A出现的次数在总订单量中的占比，计算公式：

置信度代表的是物品A出现的情况下物品B同时出现的概率，计算公式：

提升度代表的是物品A的出现使物品B出现概率提升的幅度，计算公式：

2、关联规则与协同过滤的区别

答：

（1）关联规则是基于已有订单进行的分析，是对商品之间的关系进行挖掘；而协同过滤则是基于用户偏好，是对用户行为的相似度进行挖掘。

（2）由于关联规则是基于已有订单的挖掘，所以不存在协同过滤算法中冷启动的问题。

（3）关联规则算法解决的是用户当下选择的问题，而协同过滤则是解决用户长期偏好的问题。

3、为什么我们需要多种推荐算法

答：因为每个推荐算法都有其优劣势以及适用的领域，将多种推荐算法的结果综合起来进行混合推荐才能更好以用户为中心，提供更为准确的推荐结果。

4、关联规则中的最小支持度、最小置信度该如何确定

答：在不同场景下，不同数据集最小支持度与最小置信度的现实含义不尽相同，所以没有明确的规定，而是根据不同数据集进行实验得出。一般情况下，最小支持度的范围是0.01到0.5；而最小置信度的范围是0.5到1。

5、如果通过可视化的方式探索特征之间的相关性

答：

（1）针对整体特征之间相关性的可视化，可以通过热力图的形式将特征之间的相关系数表进行可视化呈现。

（2）针对单个特征之间相关性可以通过散点图、KDE图等形式进行可视化呈现。