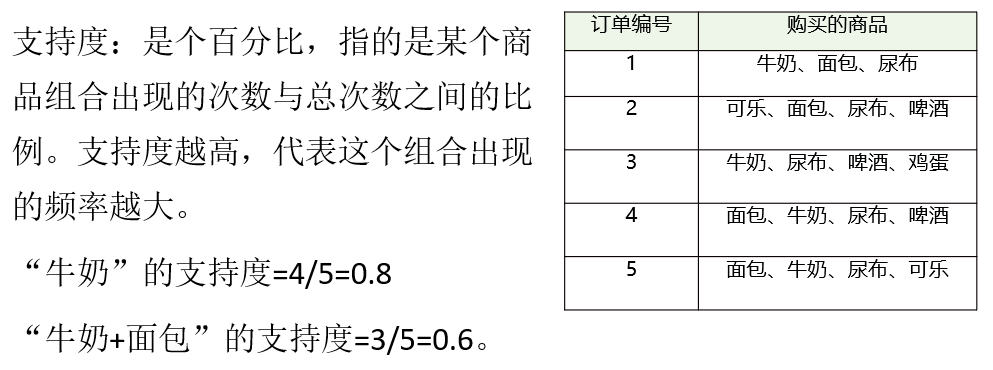
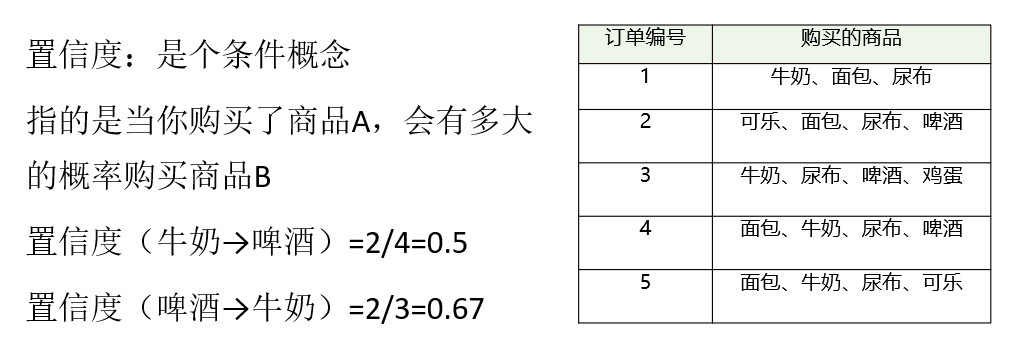
**Thinking1：关联规则中的支持度、置信度和提升度代表的什么，如何计算**

1. 支持度support：是一个百分比，指的是某个商品组合出现的次数与总次数之间的比例。支持度越高， 代表这个组合出现的频率越大



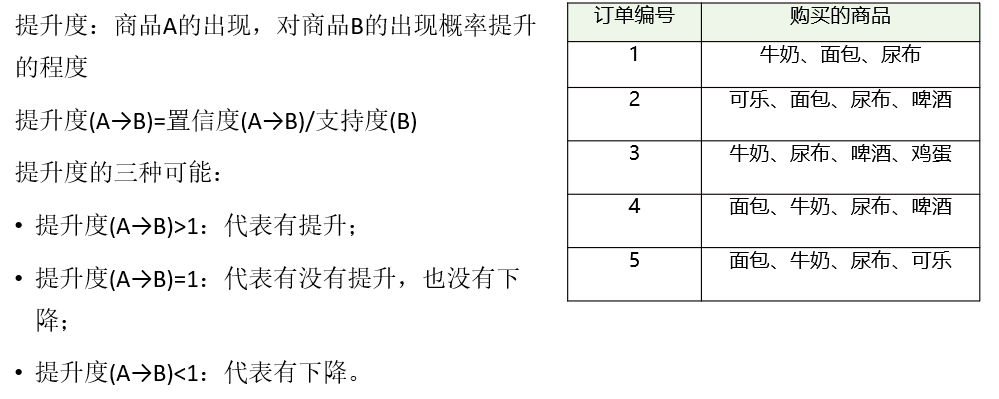
1. 置信度confidence：是个条件概念

指的是当你购买了商品A，会有多大的概率购买商品B



1. 提升度lift：商品A的出现，对商品B的出现概率提升的程度

提升度(A→B)=置信度(A→B)/支持度(B)



**Thinking2：关联规则与协同过滤的区别**

1. 协同过滤是基于用户偏好，有主观性的评价，而关联规则是基于transaction，是客观的事实；
2. 协同过滤计算的是相似度，找到相似的小群体，而关联规则基于的是全部的数据；
3. 关联规则没有利用用户偏好，而是基于购物订单进行的频繁项集挖掘；
4. 关联规则可用于当下的需求，协同过滤可反应长期的偏好。
5. 协同过滤可个性化，关联规则更多考虑整体性。

**Thinking3：为什么我们需要多种推荐算法**

不同算法代表了不同的思考维度，适用于不同的使用场景，把多种算法综合起来作为一个混合算法更加完善。

**Thinking4：关联规则中的最小支持度、最小置信度该如何确定**

通过实验来确定：

最小支持度在0.01-0.5之间，可以从高到底输出前几个项集的支持度作为参考；

最小置信度在0.5-1之间；

提升度至少要大于1.

**Thinking5：如果通过可视化的方式探索特征之间的相关性**

通过计算特征之间的person系数，绘制热力图获得特征之间的相关性。其中颜色越深代表相关性越高，颜色越浅代表相关性越低。