# Thinking1

1. 支持度=某个商品组合出现次数/总次数，表示某个商品组合出现的概率。
2. 置信度(A→B)=当你购买了A之后有多大概率购买B，是个条件概率即Pr[B|A]
3. 提升度(A→B)=置信度(A→B)/支持度(B)，表示商品A的出现，对商品B的出现概率提升的程度

# Thinking2

1. 关联规则是寻找数据项之间的相关关系，面向的是transcation，并不设计到用户偏好信息，忽略了个性化的场景。
2. 协同过滤则是先协同（群体行为）再过滤（个人行为），依赖用户偏好信息，对不同的人可以产生不同的推荐结果，即个性化。

# Thinking3

为了适应场景变化，不同的场景适用不同的推荐算法。

比如feed信息流，一般按模型得分进行倒排推荐，那么余弦相似或者XGBoost等算法比较合适。而搜索引擎查询聚类以进行流量推荐这些场景，则比较适合聚类算法。

# Thinking4

这两个阈值给得越高，得到的有相关性的item的就越好，item的数量也就越少。即item的数量和相关性成反比。最小支持度和最小置信度没有固定的取值要求，需要根据不同场景来尝试调整阈值。

# Thinking5

1. 两两特征之间可以绘制线性相关性的图来查看相关性
2. 多特征之间可以绘制散点图矩阵来查看相关性

常见绘图工具可以选择matplotlib或者seaborn