姓名： 学号：

|  |  |
| --- | --- |
| **顾客ID** | **购买商品** |
| 1 | K, A, D, B |
| 2 | D, A, C, E, B |
| 3 | C, A, B, E |
| 4 | B, A, D |

1. 给定右表所示购物篮记录数据：
2. 假设支持度阈值minsup=60%，找出所有的频繁项集；
3. 假设置信度阈值minconf=80%，仅利用置信度作为指标，在上一题频繁项集的基础上，找出所有符合如下形式的强关联规则：商品1、商品2→商品3。

解：(a) 在该记录数据中，有6个候选1项集：

{A}、{B}、{C}、{D}、{E}、{K}

对所有候选1项集计算支持度：

因此，共3个频繁1项集：

{A}、{B}、{D}

根据以上3个频繁1项集，共能组成3个候选2项集：

{A, B}、{B, D}、{A, D}

对所有候选2项集计算支持度：

因此，共3个频繁2项集：

{A, B}、{B, D}、{A, D}

根据以上3个频繁2项集，共能组成1个候选3项集：

{A, B, D}

对候选3项集计算支持度：

因此，共1个频繁3项集：

{A, B, D}

综上，所有频繁项集：{A}、{B}、{D}、{A, B}、{B, D}、{A, D}、{A, B, D}

(b) 在所有频繁项集中，符合 商品1、商品2→商品3 的规则有三个：

{A, B}→{D}

{A, D}→{B}

{B, D}→{A}

分别计算置信度：

综上，共有两个强关联规则：{A, D}→{B}，{B, D}→{A}

1. 有以下四个二维空间上的对象X1=(1, 0), X2=(0, 1), X3=(2, 1), X4=(3, 3)，利用K均值算法的一次迭代形成了两个簇C1={X1, X3}, C2={X2, X4}。那么再经过一次K均值算法的迭代，形成的两个新的簇是怎样的？注：质心为簇内对象均值，邻近度度量采用欧几里得距离。

（提示：先利用形成的两个簇确定质心，再通过邻近度将每个点分配到最近的质心）

解：设m1、m2分别为簇C1、C2的质心，则第一次迭代后：

计算每个点到m1、m2的距离：

X1属于C1

X2属于C1

X3属于C1

X4属于C2

再一次迭代后，形成的两个新的簇为C1={X1, X2，X3}, C2={X4}