



13. 给定如表 A3-1 所示的一个事务数据库,写出 Apriori 算法生成频繁项目集的过程(假设 $\text{MinSuport}=50\%$)。

表 A3-1 事务数据库示例 1

TID	Itemset
1	a,c,d,e,f
2	b,c,f
3	a,d,f
4	a,c,d,e
5	a,b,d,e,f

14. 给定如表 A3-2 所示的一个事务数据库,写出 Apriori 算法生成频繁项目集的过程(假设 $\text{MinSuport}=40\%$)。

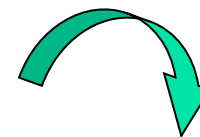
表 A3-2 事务数据库示例 2

TID	Itemset
1	1,3,4
2	2,3,4,5
3	1,3,5,7
4	2,5
5	1,2,4,6,7
6	2,4,6



- 按照步骤画出FP-Tree，并求出最大频繁集。
- 最小支持度为40%

TID	Itemset
1	a,b,c,f
2	b,c,d,e
3	a,c,e
4	b,c,d
5	b,c,d,e





KNN的例子-挖掘过程(续)

作业：假如我们统计了一些电影数据，
包括电影名称，打斗次数，接吻次数，电影类型，
如下：

电影名称	打斗次数	接吻次数	电影类型
黑客帝国	115	6	动作片
功夫	109	8	动作片
战狼	120	9	动作片
恋恋笔记本	5	78	爱情片
泰坦尼克号	6	60	爱情片
花样年华	8	69	爱情片

如果现在有一部新的电影A，它的打斗和接吻次数分别是80和7，那如何用KNN算法对其进行分类呢？



设定一组贷款数据如下， 尝试用ID3算法生成一颗树：

序号	年龄A1	是否工作A2	是否有房子A3	信贷情况A4	贷款审批结果
1	老年	否	是	很好	是
2	老年	否	是	很好	是
3	老年	是	否	好	是
4	老年	是	否	好	是
5	老年	否	否	一般	否
6	中年	否	否	一般	否
7	中年	否	否	好	否
8	中年	是	是	好	是
9	中年	否	是	很好	是
10	中年	否	是	很好	是
11	青年	否	否	一般	否
12	青年	否	否	好	否
13	青年	是	否	好	是
14	青年	是	是	一般	是
15	青年	否	否	一般	否

$$\log_2 0.333 = -1.586406$$

$$\log_2 0.667 = -0.584241$$

$$\log_2 0.2 = -2.321928$$

$$\log_2 0.8 = -0.321928$$

$$\log_2 0.4 = -1.321928$$

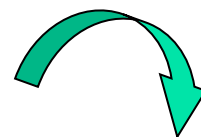
$$\log_2 0.6 = -0.736966$$



某个医院早上收了六个门诊病人，如下表：

症状	职业	疾病
打喷嚏	护士	感冒
打喷嚏	农夫	过敏
头痛	建筑工人	脑震荡
头痛	建筑工人	感冒
打喷嚏	教师	感冒
头痛	教师	脑震荡

现在又来了第七个病人，是一个打喷嚏的建筑工人。
请问他最有可能患的什么疾病？





k-means算法作业

- 根据所给的数据通过对其进行k-平均算法（设 $n=8$, $k=3$ ），进行分类。（指定1、4、7为初值中心点）

序号	属性1	属性2
1	2	10
2	2	5
3	8	4
4	5	8
5	7	5
6	6	4
7	1	2
8	4	9



AGNES算法&DIANA算法作业

- 在所给数据集上分别运行AGNES算法以及DIANA算法，假定算法的终止条件为3个簇。

序号	属性1	属性2
1	2	10
2	2	5
3	8	4
4	5	8
5	7	5
6	6	4
7	1	2
8	4	9

- 思考：两种算法的结果是否一致



- 设有12个样本点（样本点的具体坐标可以根据图中数据的位置获得，设横、纵坐标的单位间隔为1）。指定DBSCAN聚类的参数中： $r=1$ ， $\text{minPoints}=4$ 。假设有数据集如下图所示，请进行聚类。

	A	B	C	D	E	F	G	H	I	J	K	L
x	1	1	2	2	2	2	3	4	5	5	5	6
y	3	2	4	3	2	1	2	2	3	2	1	2