数据降维算法：

降低数据的维度，以便于计算

栗子：

数学 英语 语文 政治 历史 地理 生物 物理 化学 体育 美术 音乐 计算机

根据学生的成绩表，预测这个人是否聪明

在做机器学习的时候，不是所有的特征都一定有很大作用

主成成分分析（PCA）：

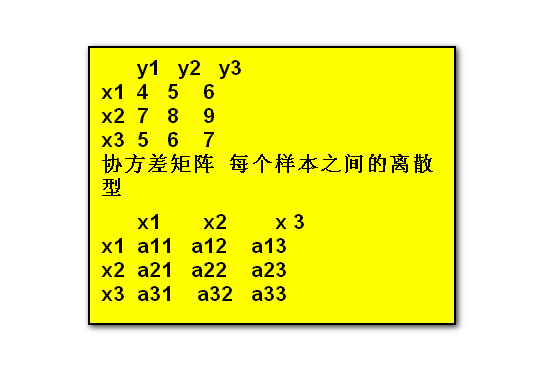
主要成分：

方差：

方差的作用：评估数据的离散程度

协方差：两个向量的离散程度

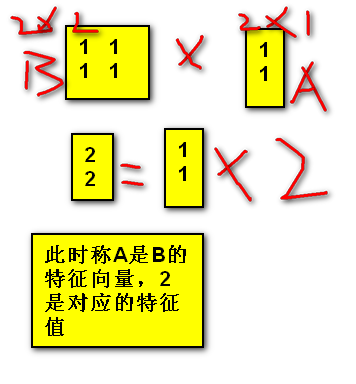
协方差矩阵：



方阵：行列都相等的矩阵

方阵特征值与特征向量：

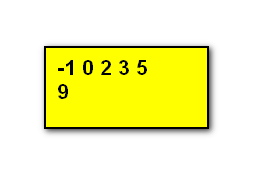
假设存在一个向量A和一个值m，使方阵B\*A=A\*m，那么称A是方阵B的特征向量，m是对应的特征值



公式：

PCA开发思路：

1. 加载数据
2. 根据特征矩阵构建协方差矩阵
3. 求协方差矩阵的特征值与特征向量
4. 将特征值加起来占比超过90%对应的特征向量提取出来构建新的特征矩阵



Numpy的科学计算

**import** numpy.linalg **as** lg *# 科学计算包***import** numpy **as** np  
*#A=np.mat('1,2,3;4,5,6')  
#A=np.mat('1 2 3 ;4 5 6') # 把字符串转为矩阵*A=np.array([[1,1],[1,1]])  
*#print(lg.eigvals(A)) # 求A的特征值  
#print(A)  
# X,Y=lg.eig(A) # 特征值与特征向量(按列对应)  
# print(X)  
#print(A.dot(Y[0:,0]))  
# print(Y[0:,0]\*X[0])  
# A\*x=B 求解x*a = np.array([[3,1], [1,2]])  
b = np.array([9, 8])  
x = lg.solve(a, b)  
print(x)

SVD 奇异值分解

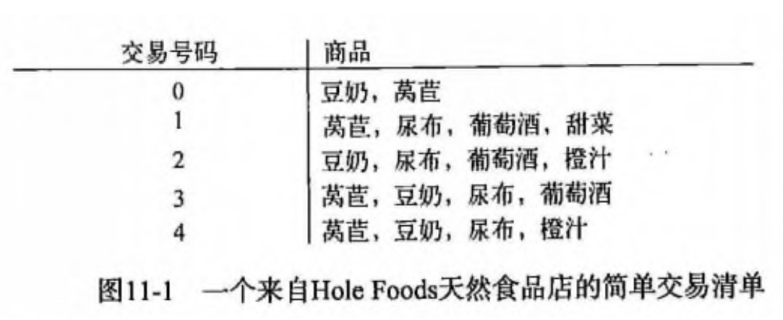
奇异值：特征值的开方

根据协方差矩阵以及奇异值构建奇异值矩阵

将奇异值中加起来占据90%以上的对应的奇异向量保存

比PCA保留的信息更多

关联性分析算法：



最小支持度：一个项集出现的最少次数

支持度: 数据集中包含该项集的记录所占的比例。例如上图中，{豆奶} 的支持度为 4/5。{豆奶, 尿布} 的支持度为 3/5

可信度: 针对一条诸如{尿布} -> {葡萄酒} 这样具体的关联规则来定义的。这条规则的可信度被定义为支持度({尿布, 葡萄酒})/支持度({尿布})，从图中可以看出支持度({尿布, 葡萄酒}) = 3/5，支持度({尿布}) = 4/5，所以{尿布} -> {葡萄酒} 的可信度 = 3/5 / 4/5 = 3/4 = 0.75

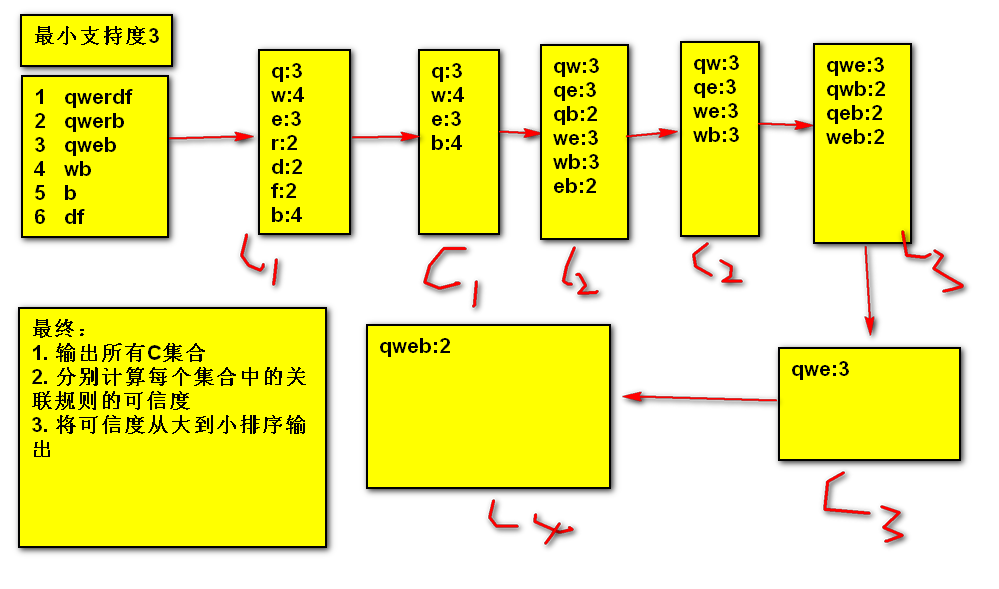
频繁项集（frequent item sets）: 经常出现在一块的物品的集合，上图中的 {葡萄酒, 尿布, 豆奶} 就是一个频繁项集的例子

关联规则: 暗示两种物品之间可能存在很强的关系，上图中尿布 -> 葡萄酒就是一个关联规则。这意味着如果顾客买了尿布，那么他很可能会买葡萄酒

Apriori原理：

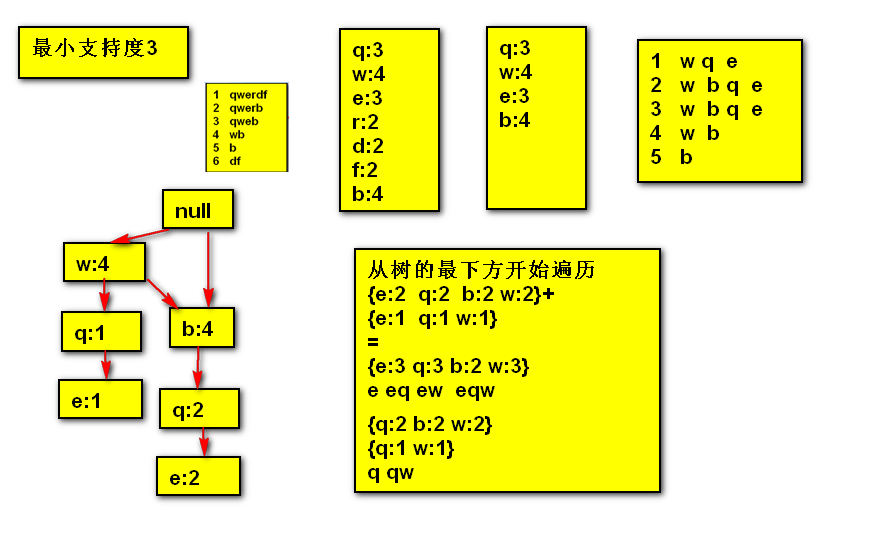
如果一个项集是频繁项集，那么它的所有子集也是频繁项集

如果一个项集是非频繁项集，那么它所有父类集合都是非频繁项集



Fp-growth树

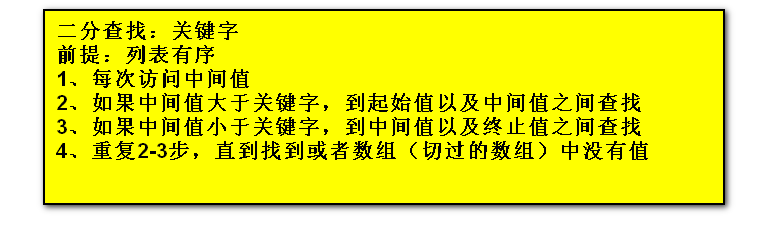
改进算法：改进查找频繁项集的思路（策略）



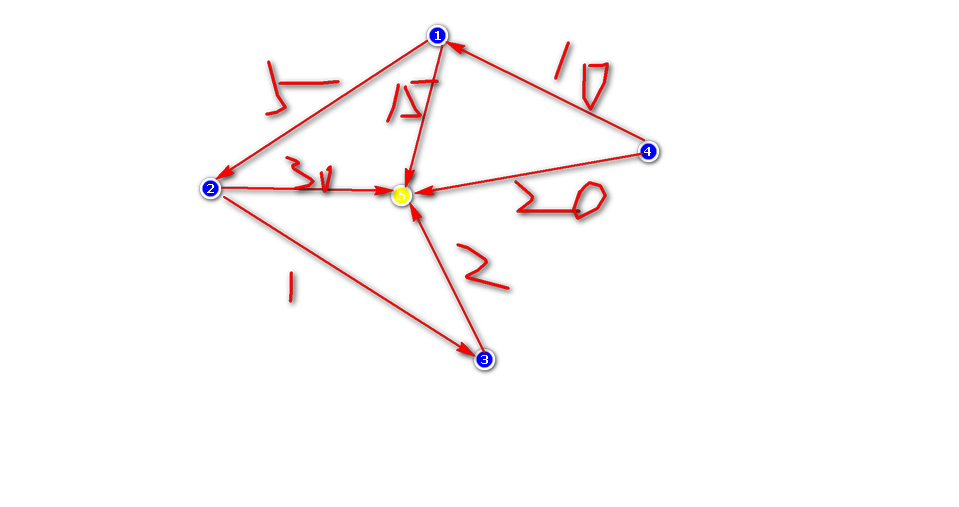
二分查找：

给定列表，查找某个数是否存在

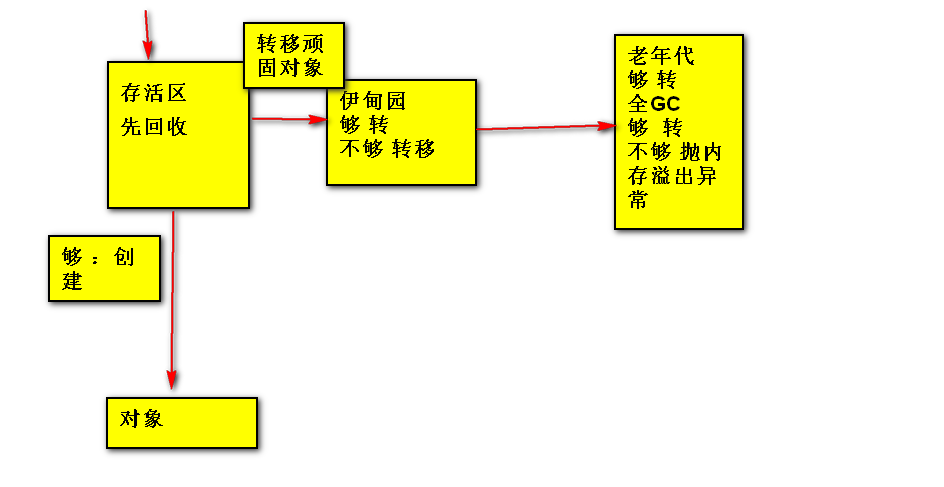
遍历查找 O(n)



最短路径算法：



分代回收



http协议的原理 底层是tcp协议

三次握手或者四次握手

