**Week6**

**###Set1\_data\_fa.csv**

**data<-read.csv("Set1\_data\_fa.csv",header = TRUE)**

**attach(data)**

**###**

**meancol<-apply(data,2,mean)**

**#表示对数据的每一列应用mean函数，meancol中每一个元素都是data中相应列的平均值**

**meancolmat<-t(replicate(dim(data)[1],meancol))#用t函数进行转置**

**data\_c<-data-meancolmat#整个data数据框进行数据中心化处理**

**covX<-cov(data\_c)#计算中心化后的协方差**

**summary(data\_c)#显示中心化后的数据信息**

**eig<-eigen(covX)#计算协方差矩阵的特征值和特征向量**

**explained\_var<-eig$values#获取特征值**

**explained\_var**

**pevar<-explained\_var/sum(explained\_var)\*100#计算每个主成分解释的方差的百分比**

**pevar**

**cumsum(pevar)#计算累积解释的方差的百分比**

**plot(1:11,pevar,type="l")#绘制每个主成分解释的方差的百分比**

**V<-eig$vectors#协方差矩阵的特征向量**

**pc<-as.matrix(data\_c)%\*%V#主成分**

**reduceddata<-pc[,1:4]#只显示前面4个主成分**

**head(reduceddata)**

**###factor analysis**

**install.packages("psych")**

**data<-read.csv("Set1\_data\_fa.csv",header = TRUE)**

**out<-psych::fa(data,nfactors=4,rotate="varimax")#进行因子分析**

**out$loading#显示因子载荷**

**out$scores#显示因子得分**

**reduceddata<-out$scores#创建只包含因子得分的新数据集**

**lambda<-out$loadings#获取因子载荷**

**psi<-diag(out$uniquenesses)#获取唯一性**

**estcov<-lambda%\*%t(lambda)+psi#计算估计的协方差矩阵**

**oricov<-cov(scale(data))#计算原始数据的协方差矩阵**

**round(oricov-estcov,5)#计算原始协方差和估计协方差矩阵的差（可以用来评估因子分析模型的拟合度）**

**#如果拟合很好，矩阵应该趋近于0矩阵**

**###apriori algorithm**

**data<-read.csv("Set1\_retail.csv",header = TRUE)**

**datacomplete<-data[complete.cases(data),]#删除所有含缺失值的行**

**head(data)**

**dim(data)#显示数据的维度**

**dim(datacomplete)#显示删除缺失值后的数据维度**

**require(tidyverse)**

**#对每个发票号码进行分组，并计算每个发票的总项目数和不同项目数**

**datacomplete%>%group\_by(invoiceno)%>%summarize(n\_total=n(),**

**n\_items=n\_distinct(stockcode))**

**install.packages("arules")**

**require(arules)**

**#根据发票号码对描述进行分组**

**split(datacomplete$description,datacomplete$invoiceno)**

**head(datacomplete)**

**#创建一个列表，其中每一个元素都是一个发票的所有项目**

**datacompletelist<-split(datacomplete$description,datacomplete$invoiceno)**

**trx<-transactions(datacompletelist)#创建一个交易数据集**

**#使用Apriori算法进行关联规则挖掘**

**ap<-apriori(trx,parameter = list(supp=0.05,conf=0.8))**

**#显示前十个关联规则**

**inspect(head(ap,10))**

**###set transaction names**

**#创建一个列表**

**a\_list<-list(c("a","b","c"),**

**c("a","b"),**

**c("a","b","d"),**

**c("c","e"),**

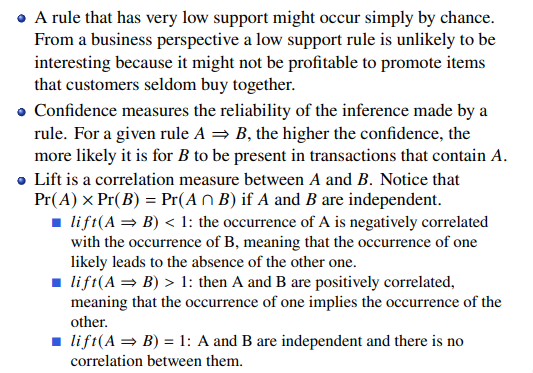
**c("a","b","d","e"))**

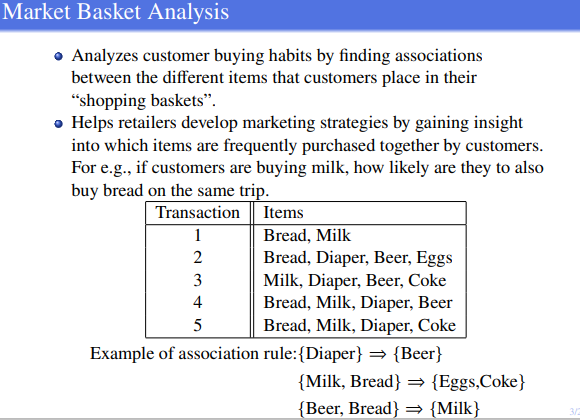
**#为列表中的每个元素设置一个名字**

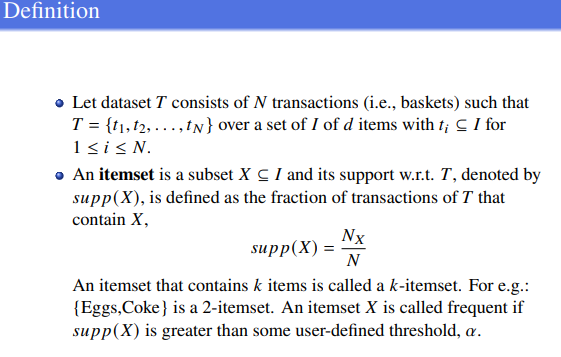
**names(a\_list)<-paste("Tr",c(1:5),sep="")**

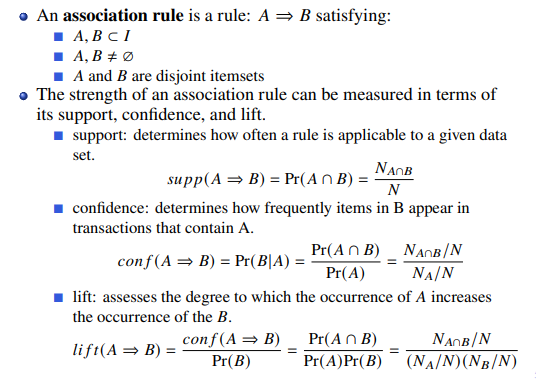
**#显示列表**

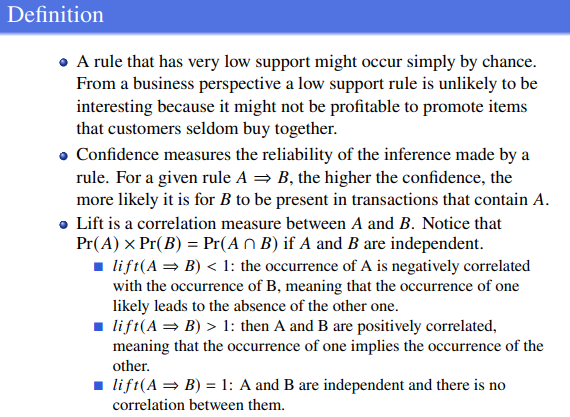
**a\_list**

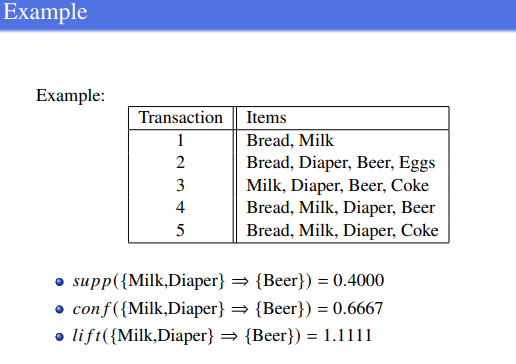
****

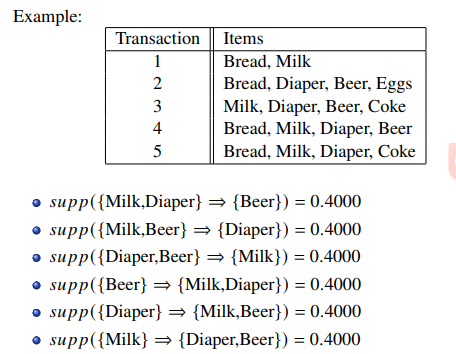
****

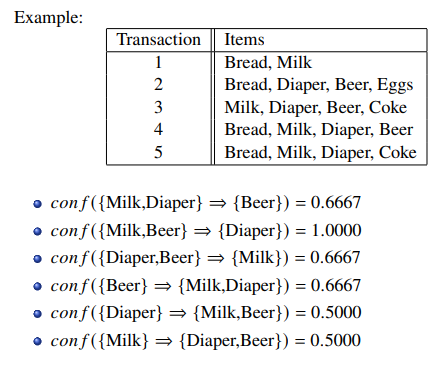
****

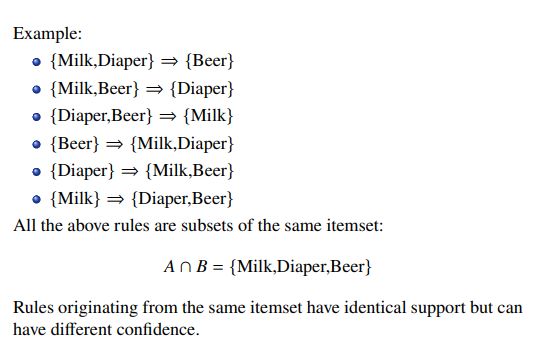
****

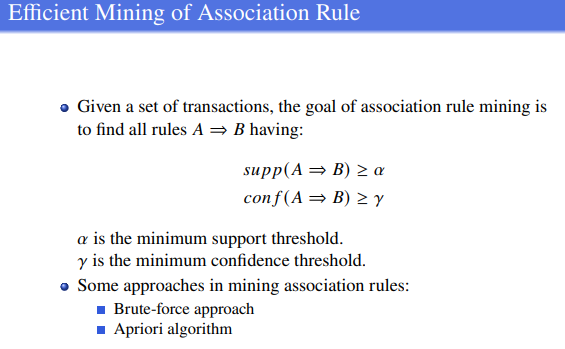
****

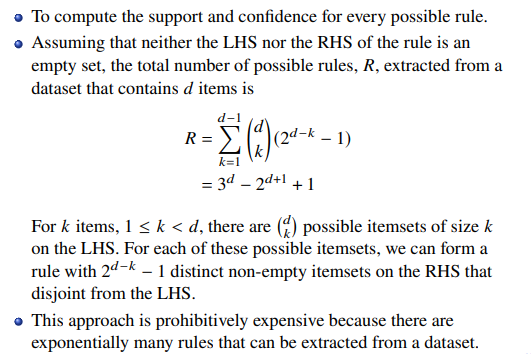
****

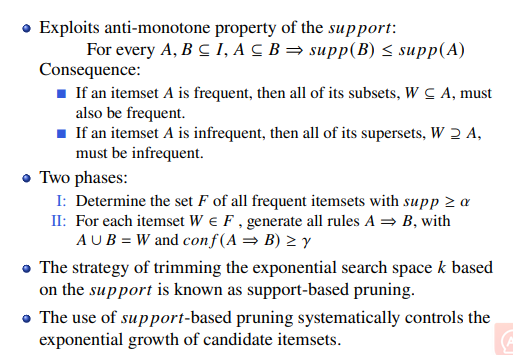
****

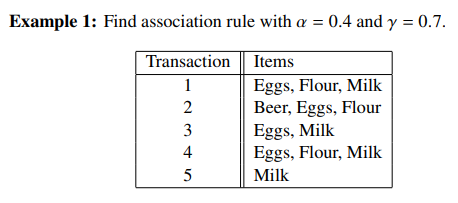
****

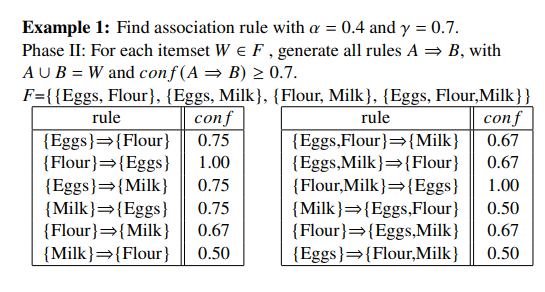
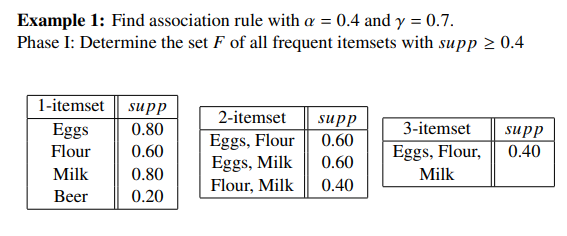
****

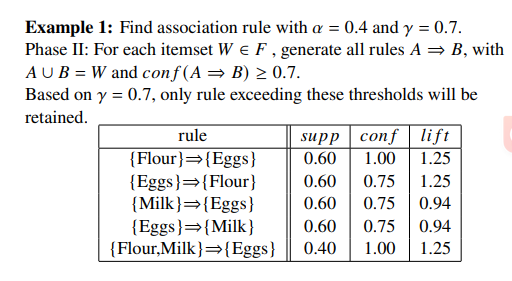
****

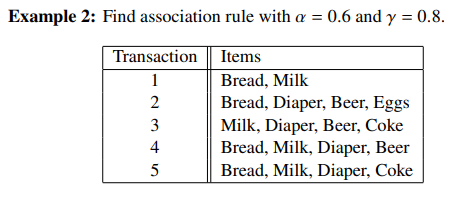
****

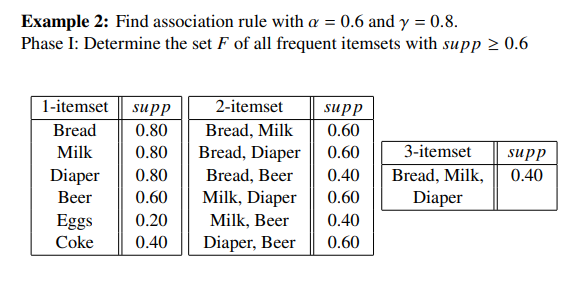
****

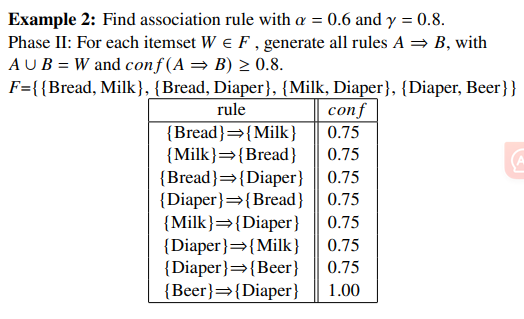
****

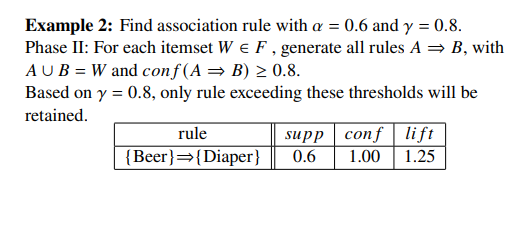
****

****

****

****

****

****