Kmeans-Davise.Bouldin-DBI指数

c1p

2016年8月30日

DBIndex<-function(x,n,m){  
 x1<-as.data.frame(x)  
 DBI<-c()  
 DIm<-c()  
 TTW<-c()  
 BET<-c()  
 TTW2<-c()  
 BET2<-c()  
 for(ij in n){  
 for(i1 in 1:m){  
 cluster1<-kmeans(x1,centers=ij)  
 a1<-split.data.frame(x1,cluster1$cluster)  
 k<-c()  
 c<-c()  
 ax<-c()  
 bx<-c()  
 Di<-c()  
 mk<-c()  
 R11<-c()  
 for(i in 1:ij){  
 ax<-as.data.frame(a1[i])  
 bx<-cluster1$centers[i,]  
 for(j in 1:nrow(ax)){  
 c<-sqrt(sum((ax[j,]-bx)^2)/nrow(ax))##在计算  
 }  
 k[i]<-sum(c)  
 }  
   
 Rij<-as.matrix(dist(cluster1$centers))##计算公式中的Rij  
 for(i in 1:ij){  
 mk<-cbind(mk,(k+k[i]))  
 }  
   
 mk1<-mk/Rij  
 for(i in 1:ij){  
 mkr<-mk1[i,]  
 R11<-c(max(mkr[which(!(is.infinite(mkr)))]),R11)  
 }  
 DIm<-c(DIm,sum(R11))  
 }  
 DBI<-c(DBI,sum(DIm)/m)  
 }  
 print(DBI)  
 par(mfrow=c(1,1))  
 plot(n,DBI,type="l",main = "DBI 指数")  
}

簇中点到该簇中心点之间的距离用的是“欧几里得距离”（可以使用别的距离计算方法） 各个中心点之间距离同样使用的是“欧几里得距离”

wiki参考 <https://en.wikipedia.org/wiki/Davies%E2%80%93Bouldin_index>

使用DBIndex函数 ##n是分类的数量，可以是一个数也可以是个连续数值

如2：10

x是输入的数据集 kmeans因为中心点初始是随机的，

设置m是重复实验多少次

a<-iris[1:4]  
DBIndex(a,2:8,20)

