Clustering a Similarity Matrix

Parismita Das

5 March 2017

Creating Similarity Matrix

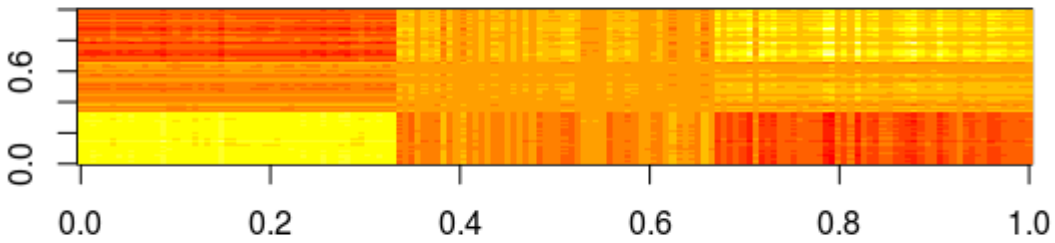
We are given with a symmetrical Dissimilarity matrix, and as all elements are positive in the given matrix, the kernel SOM and relational SOM are equivalent. Hence using the equation:

$$s(i,j) = -\frac{1}{2}\left(\delta^2(x_i,x_j) - \frac{1}{n}\sum_{k=1}^{n}\delta^2(x_i,x_k) - \frac{1}{n}\sum_{k=1}^{n}\delta^2(x_k,x_j) + \frac{1}{n^2}\sum_{k,k'=1}^{n}\delta^2(x_k,x_{k'})\right)$$

The algorithm to calculate the Similarity Matrix

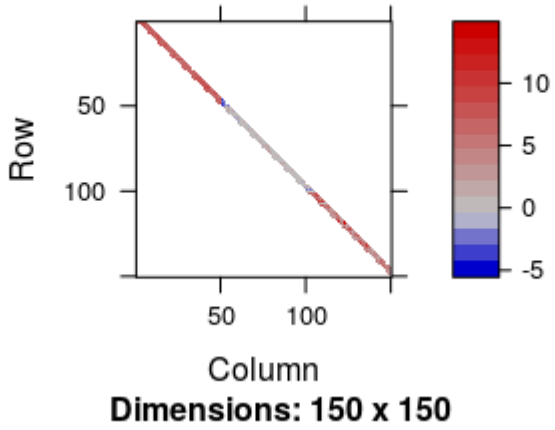
```
# algo to calculate similarity matrix s from the equation given where
#K is the dissimilatriy matrix delta(xi,xj)
#Dkk is sum[k,k':1 to n](delta(xk,xk')^2), Dij is delta(xi,xj)^2,
#Dik and Dkj is sum[k:1 to n](delta(xi,xk)^2) and sum[k:1 to n](delta(xk,xj)^2) resp
n <- 150
Dkk <- sum(K^2)
s <- c()
for (i in 1:n) {
  Dik <- sum(K[i,]^2)
  for (j in 1:n) {
    Dkj <- sum(K[,j]^2)
    Dij <- K[i,j]^2
    #the equation
    sij <- -0.5*(Dij - 1/n*(Dik+Dkj) + 1/n^2*Dkk)
    s <- c(s,sij)
    }
}

# the similarity matrix
s <- matrix(s , nrow = 150,ncol = 150)
```

We get the Similarity Matrix Image as:



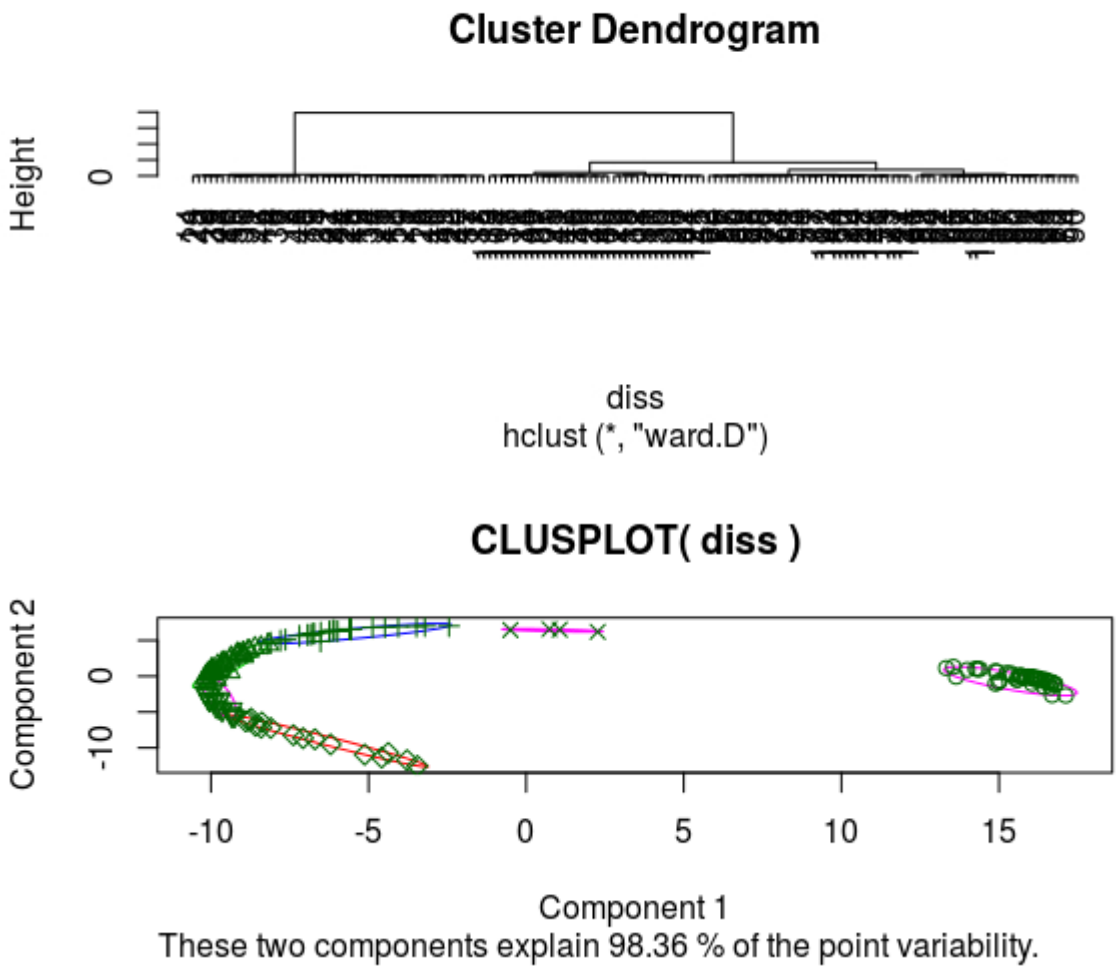
Now Extracting the diagonal band of width 5 that is to be used for clustering, using the function HeapHop

```
# extracting diagonal band
low <- 1
high <- h
delta <- col(s) - row(s)
s[delta < low | delta > high] <- 0
```



Creating Dendrograms and cluster plot, clustering via hclust

```
diss <- dist(K)
hc = hclust(d = diss, method = 'ward.D')
y_hc = cutree(hc,6)
plot(hc)
diss=as.matrix(diss)
clusplot(diss, y_hc, lines = 0, color = TRUE)
```



We can see that only 3 major clusters are formed