相似系数与距离之间的关系

距离⇒相似系数

$$C_{ij}=\frac{1}{1+d_{ij}}$$

相似系数⇒距离 相似系数矩阵$(C_{ij})$非负定.

$$d_{ij}=\sqrt{2(1-C_{ij})}$$

dist (距离矩阵) → hclust (层次聚类)

(WGCNA也是层次聚类，聚gene module).

MDS (多维标度法)

核心目的：从距离矩阵D，在低维空间中寻找一个矩阵Z，使得Z构成的欧式距离矩阵B尽可能和原来矩阵保持一致。此种多维标度法也称作主座标分析。

① 样本矩阵 $X=\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm}\end{bmatrix}=[\vec{x}_1,\vec{x}_2,\cdots\vec{x}_m]$ m个sample, n个gene

② $D=\begin{bmatrix} d(x_1,x_1) & \cdots & d(x_1,x_m) \\ \vdots & & \\ d(x_m,x_1) & & d(x_m,x_m)\end{bmatrix}$ 距离矩阵

③ $\vec{x}_i=\begin{bmatrix}x_{1i} \\ \vdots\end{bmatrix}$ $\vec{Y}_i=\begin{bmatrix}Y_{1i} \\ \vdots\end{bmatrix}$ $\vec{Y}_i$ ⇒ 新的距离矩阵B ?

④ 找到 $Y_i=\begin{bmatrix}Y_{1i} \\ \vdots \\ Y_{pi}\end{bmatrix}$，令 $B=Y^TY=\begin{bmatrix}Y_1 \\ Y_2 \\ \vdots \\ Y_m\end{bmatrix}[Y_1,Y_2,\cdots Y_m]$

中心化 (B是为了方便计算)

⑤ $dist_{ij}=d(x_i,x_j)=\|\vec{Y}_i-\vec{Y}_j\|$ (欧氏距离)

$$dist_{ij}^2=\|Y_i\|^2+\|Y_j\|^2-2\vec{Y}_i\cdot\vec{Y}_j$$

$$dist_{ij}^2=b_{ii}+b_{jj}-2b_{ij}$$

$$b_{ij}=\frac{1}{2}(b_{ii}+b_{jj}-dist_{ij})$$

⑥ $\sum_{i=1}^m b_{ij}=\sum_{i=1}^m \vec{Y}_i\cdot\vec{Y}_j=\sum_{i=1}^m=\sum_{i=1}^m(y_{1i}y_{1j}+y_{2i}y_{2j}+\cdots+y_{pi}y_{pj})$

$$=y_{11}y_{1j}+y_{21}y_{2j}+\cdots+y_{p1}y_{pj}+$$
$$y_{12}y_{1j}+y_{22}y_{2j}+\cdots+y_{p2}y_{pj}+$$
$$\vdots$$
$$y_{1m}y_{1j}+y_{2m}y_{2j}+\cdots+y_{pm}y_{pj}$$

$$=y_{1j}\underbrace{\sum_{i=1}^m y_{1i}}_{0}+y_{2j}\underbrace{\sum_{i=1}^m y_{2i}}_{0}+\cdots+y_{pj}\underbrace{\sum_{i=1}^m y_{pi}}_{0}$$

$$=0$$

⑦ $\sum_{i=1}^m dist_{ij}^2=\sum_{i=1}^m(b_{ii}+b_{jj}-2b_{ij})=\sum_{i=1}^m b_{ii}+\sum_{i=1}^m b_{jj}-2\sum_{i=1}^m b_{ij}$

$$=tr(B)+mb_{jj}+0$$

$$\sum_{j=1}^m dist_{ij}^2=tr(B)+mb_{ii}$$

$$\sum^m\sum^m \cdots=\sum^m(tr(B)+mb_{ii})$$