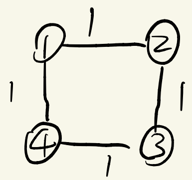
Problem 4

1. 欧式距离矩阵 $D=\begin{pmatrix} 0 & 1 & \sqrt{2} & 1 \\ 1 & 0 & 1 & \sqrt{2} \\ \sqrt{2} & 1 & 0 & 1 \\ 1 & \sqrt{2} & 1 & 0 \end{pmatrix}$

2. 邻域图




最短路径距离矩阵：$\begin{pmatrix} 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \end{pmatrix} = D_1$

3. 双中心化矩阵 $B = -\frac{1}{2} J D_1^{(2)} J$

其中 $J = I_n - \frac{1}{n} 11^T$

Problem 5.

PCA的步骤：① 计算协方差矩阵 $C = \frac{1}{n} X^T X$ ② 对C进行特征分解：$C = V \Lambda V^T$，其中 $\Lambda$ 是对角矩阵，对角线元素为特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$. $V$ 是特征向量矩阵 ③ 选择前k个特征向量 $V_k \in R^{d \times k}$ ④ $Z = X V_k$

经典MDS步骤：① 计算距离矩阵 $D_{ij} = \|x_i - x_j\|_2$

② 计算双中心化矩阵 $B = -\frac{1}{2} J D^2 J$ ③ 对B进行特征分解：

$B = U\Gamma U^T$，选择前k个最大特征值对应的特征向量

$U_k$ ④ $Y = U_k \Gamma_k^{1/2}$

因此，我们需要证明 $Y = Z$

由于 $D_{ij}^2 = \|x_i - x_j\|_2^2 = \|x_i\|^2 + \|x_j\|^2 - 2x_i^T x_j$

$J = I_n - \frac{1}{n} 1 1^T$

由于 $B = -\frac{1}{2} J D^2 J$，则：

$$B = -\frac{1}{2}\left(D^2 - \frac{1}{n} D^2 1 1^T - \frac{1}{n} 1 1^T D^2 + \frac{1}{n^2} 1 1^T D^2 1 1^T\right)$$

于是 $B_{ij} = -\frac{1}{2}\left(D_{ij}^2 - \frac{1}{n}\sum_{k=1}^{n} D_{ik}^2 - \frac{1}{n}\sum_{k=1}^{n} D_{kj}^2 + \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n} D_{kl}^2\right)$

代入 $D_{ij}^2$，则 $B_{ij} = x_i^T x_j$

于是 $C = \frac{1}{n} X^T X \qquad B = X X^T$

由线性代数知识有：

设 $X = U_左 S V_右^T$ 是奇异值分解.

$XX^T = U_左 S V_右^T V S^T U_左^T = U_左 S S^T U_左^T$

$\frac{1}{n} X^T X = \frac{1}{n} V_右 S^T U_左^T U_左 S V_右^T = \frac{1}{n} V_右 S^T S V_右^T$

故B的特征向量为 $U_左$　C的特征向量为 $V_右$

且 $SS^T = \Gamma$　　S为对角奇异值矩阵

故 $Z = X V_右 = U_左 S V_右^T V_右 = U_左 S$

$$Y = U_{左} \Gamma^{\frac{1}{2}} = U_{左} S = Z.$$