COMP8118-A3

Monday, September 19, 2022 5:11 PM

Q1

2dimensional vectors:	(2, 2)	(4, 4)	(1, 5)	(5, 1)

L = 2, k = 1

Step 1

Mean vector = ((2 + 4 + 1 + 5)/4, (2 + 4 + 5 + 1)/4 = (3, 3)

Differences	=	(-1, -1)	(1, 1)	(-2, 2)	(2, -2)

Step 2:

$$Y = \begin{bmatrix} -1 & 1 & -2 & 2 \\ -1 & 1 & 2 & -2 \end{bmatrix}$$

$$\Sigma = \frac{1}{4}\mathrm{YY}^{\mathrm{T}} = \frac{1}{4}\begin{bmatrix} -1 & 1 & -2 & 2 \\ -1 & 1 & 2 & -2 \end{bmatrix}\begin{bmatrix} -1 & -1 \\ 1 & 1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 10 & -6 \\ -6 & 8 \end{bmatrix} = \begin{bmatrix} 2.5 & -1.5 \\ -1.5 & 2.5 \end{bmatrix}$$

Step 3:

Eigen Values: $\begin{vmatrix} 2.5-\lambda & -1.5 \\ -1.5 & 2.5-\lambda \end{vmatrix} = 0 \rightarrow (2.5 - \lambda)^2 - (-1.5)^2 = 0$

$\rightarrow \lambda^2 - 5\lambda + 4 = 0$

$\rightarrow (\lambda - 4)(\lambda - 1) = 0$

$$\lambda_1 = 4: eigen\ vector\ 1 = \left[\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}\right], \lambda_2 = 1: eigen\ vector\ 2 = \left[\sqrt{\frac{1}{2}}, -\sqrt{\frac{1}{2}}\right]$$

Step 4:

$$\Phi = \begin{bmatrix} \sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}}, -\sqrt{\frac{1}{2}} \end{bmatrix}$$

Step 5: K-L transform

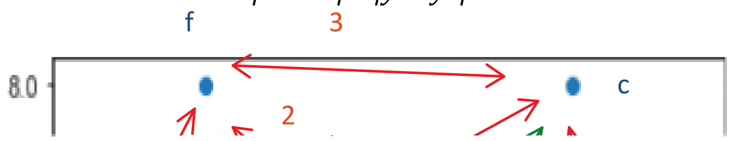
$$Y = \Phi^T X = \begin{bmatrix} \sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}}, -\sqrt{\frac{1}{2}} \end{bmatrix}\begin{bmatrix} 2 & 4 & 1 & 5 \\ 2 & 4 & 5 & 1 \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} & 4\sqrt{2} & 3\sqrt{2} & 3\sqrt{2} \\ 0 & 0 & -2\sqrt{2} & 2\sqrt{2} \end{bmatrix}$$

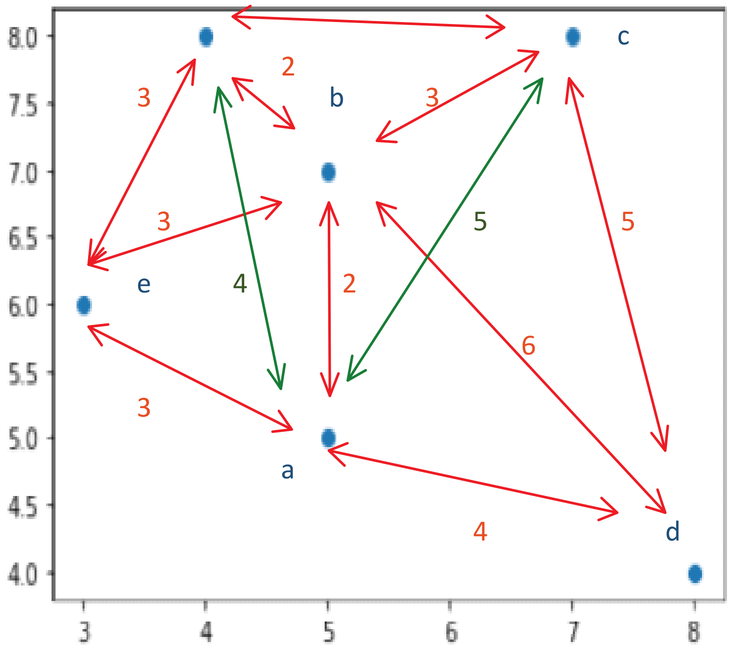
$\rightarrow subspace: \left[0, 0, -2\sqrt{2}, 2\sqrt{2}\right]$

Q2

vectors:	(5, 5)	(5, 7)	(7, 8)	(8, 4)	(3, 6)	(4, 8)
names:	a	b	c	d	e	f

Manhattan distance: |x1 - x2| + |y1 - y2|





a. *Distance-based*

$\varepsilon = 4$
N0 = 3

N	*4*	*4*	*2*	*1*	*3*	*4*
names	*a*	*b*	*c*	*d*	*e*	*F*

N(d) < N(c) < N(e) = N0 = 3 -> c, d, e are outliers given the conditions

b. *Density-based*

K = 3

3-distance(c) : distance(c, d) = distance(c, a) = 5
3-distance(d) : distance(d, b) = 6
3-neighborhood (c) = {a, b, d, f} -> $\varepsilon_c = 5$
3-neighborhood (d) = {a, b, c} -> $\varepsilon_d = 6$

Lrd3(c) = 1/ 5 = 0.2
Lrd3(c) = 1/ 6 = 0.1667

LOF(c) = $\frac{\sum\frac{lrd3(o)}{lrd3(c)}}{3} = \frac{lrd3(f)+lrd3(b)+lrd3(a)+lrd3(d)}{\frac{3}{5}} = \frac{\frac{1}{3}+\frac{1}{3}+\frac{1}{4}+\frac{1}{6}}{\frac{3}{5}} = 1.25$

LOF(d) = $\frac{\sum\frac{lrd3(o)}{lrd3(d)}}{3} = \frac{lrd3(a)+lrd3(b)+lrd3(c)}{\frac{3}{5}} = \frac{\frac{1}{4}+\frac{1}{3}+\frac{1}{5}}{\frac{3}{6}} = 1.5666$

1.56 > 1.25: d is a local outlier