تمرین سری 8 درس ML - رادین خیام - 99101579

سؤال 1 -

الف)

$$d_1(\underline{x}, \underline{y}) = \sum_{k=1}^{2} |x_k - y_k|$$

$D_1 = \{x_1, x_2, x_3, x_4\}$, $x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $x_2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$, $x_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$, $x_4 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \longrightarrow m_{D_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$D_2 = \{y_1, y_2, y_3, y_4\}$, $y_1 = \begin{bmatrix} -3 \\ 0 \end{bmatrix}$, $y_2 = \begin{bmatrix} -5 \\ 0 \end{bmatrix}$, $y_3 = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$, $y_4 = \begin{bmatrix} -5 \\ -2 \end{bmatrix} \longrightarrow m_{D_2} = \begin{bmatrix} -4 \\ -1 \end{bmatrix}$

$d_1(x_1, y_1) = 3$, $d_1(x_1, y_2) = 5$, $d_1(x_1, y_3) = 5$ .....

$$A_{ij} = d_1(x_i, y_j) \longrightarrow A = \begin{bmatrix} 3 & 5 & 5 & 7 \\ 5 & 7 & 7 & 9 \\ 5 & 7 & 7 & 9 \\ 7 & 9 & 9 & 11 \end{bmatrix} \longrightarrow$$

$m_{D_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $m_{D_2} = \begin{bmatrix} -4 \\ -1 \end{bmatrix}$

$d_{max} = 11$

$d_{min} = 3$

$d_{avg} = 7$

$d_{mean} = d_1\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -4 \\ -1 \end{bmatrix}\right) = 7$

$d_e = \sqrt{\frac{4 \times 4}{4+4}}\, d_{mean} = 7\sqrt{2}$

ب)

$$d_\infty(\underline{x}, \underline{y}) = \max |x_k - y_k|$$

$\longrightarrow d_\infty(x_1, y_1) = 3$, $d_\infty(x_1, y_2) = 5$, ...

$$A_{ij} = d_\infty(x_i, y_j) \longrightarrow A = \begin{bmatrix} 3 & 5 & 3 & 5 \\ 3 & 5 & 4 & 5 \\ 5 & 7 & 5 & 7 \\ 5 & 7 & 5 & 7 \end{bmatrix} \longrightarrow$$

$d_{max} = 7$

$d_{min} = 3$

$d_{avg} = 5.06$

$d_{mean} = d_\infty\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -4 \\ -1 \end{bmatrix}\right) = 5$

$d_e = \sqrt{\frac{4 \times 4}{4+4}}\, d_{mean} = 5\sqrt{2}$

سؤال 2 -

الف)

$C = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$

$x_1 = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix}, x_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, x_3 = \begin{bmatrix} 2.5 \\ 1.75 \end{bmatrix}, x_4 = \begin{bmatrix} 1.5 \\ 2 \end{bmatrix}, x_5 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, x_6 = \begin{bmatrix} 1 \\ 3.5 \end{bmatrix}, x_7 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

$x_8 = \begin{bmatrix} 3.5 \\ 3 \end{bmatrix}, \quad x = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$

$d_1(x, x_1) = 7, \; d_1(x, x_2) = 7, \; d_1(x, x_3) = 5.75, \; d_1(x, x_4) = 6.5$

$d_1(x, x_5) = 5, \; d_1(x, x_6) = 5.5, \; d_1(x, x_7) = 5, \; d_1(x, x_8) = 3.5$

$\rightarrow \boxed{d_{min} = 3.5}, \; \boxed{d_{max} = 5}, \; \boxed{d_{avg} = \frac{45.25}{8} = 5.66}$

ب)

$d_\infty(x, x_1) = 4.5, \; d_\infty(x, x_2) = 4, \; d_\infty(x, x_3) = 3.5$

$d_\infty(x, x_4) = 4.5, \; d_\infty(x, x_5) = 3, \; d_\infty(x, x_6) = 5, \; d_\infty(x, x_7) = 4, \; d_\infty(x, x_8) = 2.5$

$\rightarrow \boxed{d_{min} = 2.5}, \; \boxed{d_{max} = 5}, \; \boxed{d_{avg} = \frac{31}{8} = 3.9}$

سؤال 3-

$$s(x,y) > 0 \quad \forall x,y \in X \quad , \quad d(x,y) = \frac{a}{s(x,y)} \quad a>0$$

فرض می کنیم که s_0 ماکسیمم مقدار $s(x,y)$ است.

$$\rightarrow -\infty < s(x,y) < s_0 < +\infty \rightarrow -\infty < \frac{a}{s_0} \leq d(x,y) < \infty$$

$$\rightarrow \boxed{\forall x,y \in X, \exists d_0 \in \mathbb{R} \quad -\infty < d_0 \leq d(x,y) < \infty}$$

شرط اول

$$\forall x \in X: s(x,x) = s_0 \rightarrow \boxed{\forall x \in X: d(x,x) = \frac{a}{s_0} = d_0}$$

شرط دوم

شرط سوم

$$s(x,y) = s(y,x) \rightarrow \frac{a}{s(x,y)} = \frac{a}{s(y,x)} \rightarrow \boxed{\forall x,y \in X: d(x,y) = d(y,x)}$$

$$s(x,y) = s_0 \leftrightarrow x=y \rightarrow \frac{a}{s(x,y)} = \frac{a}{s_0} = d_0 \leftrightarrow x=y$$

$$\rightarrow \boxed{\forall x \in X: d(x,x) = d_0} \rightarrow$$

شرط چهارم

$$\forall x,y,z \in X: \frac{1}{s(x,z)} \leq \frac{1}{s(x,y)} + \frac{1}{s(y,z)} \xrightarrow[a>0]{\times a} \frac{a}{s(x,z)} \leq \frac{a}{s(x,y)} + \frac{a}{s(y,z)}$$

$$\rightarrow \boxed{\forall x,y,z \in X: \quad d(x,z) \leq d(x,y) + d(y,z)} \rightarrow$$

شرط پنجم

بنابراین d هم شروط measure و هم metric را برای حالت عدم تشابه دارد.

سؤال 4-

$x_1=\begin{bmatrix}1\\1\end{bmatrix}, x_2=\begin{bmatrix}1\\2\end{bmatrix}, x_3=\begin{bmatrix}2\\2\end{bmatrix}, x_4=\begin{bmatrix}2\\3\end{bmatrix}, x_5=\begin{bmatrix}3\\3\end{bmatrix}, x_6=\begin{bmatrix}3\\4\end{bmatrix}$

$x_7=\begin{bmatrix}4\\4\end{bmatrix}, x_8=\begin{bmatrix}4\\5\end{bmatrix}, x_9=\begin{bmatrix}5\\5\end{bmatrix}, x_{10}=\begin{bmatrix}5\\6\end{bmatrix}, x_{11}=\begin{bmatrix}-4\\5\end{bmatrix}, x_{12}=\begin{bmatrix}-3\\5\end{bmatrix}$

$x_{13}=\begin{bmatrix}-4\\4\end{bmatrix}, x_{14}=\begin{bmatrix}-3\\4\end{bmatrix}$, $\Theta=\sqrt{2}$, Euclidean distance

BSAS: الف)

$m=1$

$C_1=\{x_1\}$

$d(x_2,C_1)=d(x_1,x_2)=1<\sqrt{2} \rightarrow C_1=\{x_1,x_2\} \rightarrow m_1=\begin{bmatrix}1\\ \frac{3}{2}\end{bmatrix}$

$d(x_3,C_1)=d(x_3,m_1)=\sqrt{\frac{5}{4}}<\sqrt{2} \rightarrow C_1=\{x_1,x_2,x_3\} \rightarrow m_1=\begin{bmatrix}\frac{4}{3}\\ \frac{5}{3}\end{bmatrix}$

$d(x_4,C_1)=d(x_4,m_1)=\sqrt{\frac{4}{9}+\frac{16}{9}}>\sqrt{2} \rightarrow C_2=\{x_4\}$

$d(x_5,C_1)=d(x_5,m_1)=\sqrt{\frac{25}{9}+\frac{16}{9}}$

$d(x_5,C_2)=d(x_5,x_4)=1$ $\xrightarrow{\min} d(x_5,C_2)=1<\sqrt{2} \rightarrow C_2=\{x_4,x_5\}$ $\nearrow m_2=\begin{bmatrix}2.5\\3\end{bmatrix}$

$d(x_6,C_1)=\sqrt{\frac{25}{9}+\frac{49}{9}}$

$d(x_6,C_2)=\sqrt{\frac{1}{4}+1}$ $\xrightarrow{\min} d(x_6,C_2)<\sqrt{2} \rightarrow C_2=\{x_4,x_5,x_6\} \rightarrow m_3=\begin{bmatrix}\frac{8}{3}\\ \frac{10}{3}\end{bmatrix}$

---

* به همین ترتیب اگر مراحل را جلو ببریم به نتیجه زیر می رسیم:

$C_1=\{x_1,x_2,x_3\}, C_2=\{x_4,x_5,x_6\}, C_3=\{x_7,x_8,x_9\}, C_4=\{x_{10}\}$

$C_5=\{x_{11},x_{12},x_{13},x_{14}\}$

MBSAS: ادامه سؤال 4 الف)

* Cluste determination:

$C_1 = \{x_1\}$

$d(x_2, C_1) = d(x_2, x_1) = 1 < \sqrt{2}$

$d(x_3, C_1) = d(x_3, x_1) = \sqrt{2} = \sqrt{2}$

$d(x_4, C_1) = d(x_4, x_1) = \sqrt{5} > \sqrt{2} \longrightarrow C_2 = \{x_4\}$

$$\begin{array}{l} d(x_5, C_1) = d(x_5, x_1) \\ d(x_5, C_2) = d(x_5, x_4) \end{array} \xrightarrow{\min} d(x_5, x_4) = 1 < \sqrt{2}$$

$$\begin{array}{l} d(x_6, C_1) = d(x_6, x_1) \\ d(x_6, C_2) = d(x_6, x_4) \end{array} \xrightarrow{\min} d(x_6, x_4) = \sqrt{2} = \sqrt{2}$$

$$\begin{array}{l} d(x_7, C_1) = d(x_7, x_1) \\ d(x_7, C_2) = d(x_7, x_4) \end{array} \xrightarrow{\min} d(x_7, x_4) = \sqrt{5} > \sqrt{2} \longrightarrow C_3 = \{x_7\}$$

* به همین ترتیب که این مراحل را جلو ببریم :

$C_1 = \{x_1\}$, $C_2 = \{x_4\}$, $C_3 = \{x_7\}$, $C_4 = \{x_{10}\}$, $C_5 = \{x_{11}\}$

حال مرحله بعد الگوریتم شروع می‌شود.

ادامه سؤال 4 الف)

$C_1 = \{x_1\}$, $C_2 = \{x_4\}$, $C_3 = \{x_7\}$, $C_4 = \{x_{10}\}$, $C_5 = \{x_{11}\}$

* Patter Classification:

$$\min_{1 \leq j \leq 5} d(x_2, C_j) = d(x_2, C_1) \rightarrow C_1 = \{x_1, x_2\} \rightarrow m_1 = \begin{bmatrix} 1 \\ \frac{3}{2} \end{bmatrix}$$

$$\min_{1 \leq j \leq 5} d(x_3, C_j) = d(x_3, C_2) \rightarrow C_2 = \{x_4, x_3\} \rightarrow m_2 = \begin{bmatrix} 2 \\ \frac{5}{2} \end{bmatrix}$$

به همین ترتیب اگر مراحل را جلو برویم:

$C_1 = \{x_1, x_2\}$, $C_2 = \{x_3, x_4, x_5\}$, $C_3 = \{x_6, x_7, x_8\}$, $C_4 = \{x_9, x_{10}\}$

$C_5 = \{x_{11}, x_{12}, x_{13}, x_{14}\}$

به صورت کلی هر دو الگوریتم نسبت به ترتیب داده‌ها حساس هستند اما الگوریتم MBSAS حساسیت کمتری دارد.

* قسمت های دیگر سؤال را گفتند که درباره‌ش بحث کنیم، خیلی متوجه نیستم باید چی بگم الان ولی خب ترتیب رو که عوض می کنیم احتمال بسیار بالا نتایج خوشه بندی تغییر می کند.

سؤال ۵ .

تغییراتی که باید در الگوریتم ها بدهیم به این صورت است :

$$d(x_i, c_k) = \min_{1 \le j \le m} d(x_i, c_j) \longrightarrow s(x_i, c_k) = \max_{1 \le j \le m} s(x_i, c_j)$$

و زمانی هم که می خواهیم با آستانه چک کنیم :

$$d(x_i, c_k) < \theta_1 \longrightarrow s(x_i, c_k) > \theta_1$$

$$d(x_i, c_k) > \theta_2 \longrightarrow s(x_i, c_k) < \theta_2$$

سؤال 6 -

الف) $q = 15 \quad , \theta = 3$

$C_1 = \{x_1\}$

$\underset{x_i}{\arg\max}\left\{\min_{C_k} d(x_i, C_k)\right\} = d(x_{15}, C_1) > 3 \longrightarrow C_2 = \{x_{15}\}$

$\underset{x_i}{\arg\max}\left\{\min_{C_k} d(x_i, C_k)\right\} = d(x_3, C_1) > 3 \longrightarrow C_3 = \{x_3\}$

$\underset{x_i}{\arg\max}\left\{\min_{C_k} d(x_i, C_k)\right\} = d(x_{13}, C_2) > 3 \longrightarrow C_4 = \{x_{13}\}$

$\underset{x_i}{\arg\max}\left\{\min_{C_k} d(x_i, C_k)\right\} = d(x_8, C_1 \text{ or } C_2 \text{ or } C_3 \text{ or } C_4) = \sqrt{13} > 3 \longrightarrow C_5 = \{x_8\}$

دیگه خوشه جدیدی پیدا نمی شود.

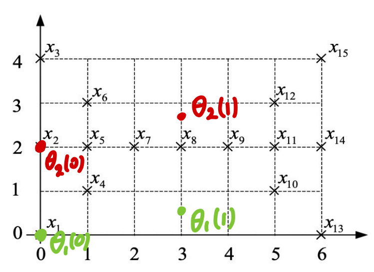
حال داده های که خوشه ندارند را به نزدیکترین خوشه می دهیم:

$\longrightarrow \quad C_1 = \{x_1, x_4, x_2\} \quad , \quad C_2 = \{x_{12}, x_{14}, x_{15}\}$

$C_3 = \{x_3, x_6\} \quad , \quad C_4 = \{x_{13}, x_{10}\} \quad , \quad C_5 = \{x_8, x_5, x_7, x_9, x_{11}\}$

ادا- سؤال 6-

ب) فرض اولیه برای مراکز خوشه‌ها:



$$\underline{\theta}_1 = x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} , \quad \underline{\theta}_2 = x_2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

$u_{11}=1$, $u_{22}=1$, $u_{32}=1$, $u_{41}=1$, $u_{52}=1$, $u_{62}=1$, $u_{72}=1$, $u_{82}=1$, $u_{92}=1$,

$u_{10,1}=1$, $u_{11,2}=1$, $u_{12,2}=1$, $u_{13,1}=1$, $u_{14,2}=1$, $u_{15,2}=1$ بقیه u_{ij} ها هم صفراند.

$$\underline{\theta}_1 = \frac{1}{4}(x_1+x_4+x_{10}+x_{13}) = \begin{bmatrix} 3 \\ \frac{1}{2} \end{bmatrix}, \underline{\theta}_2 = \frac{1}{11}(x_2+x_3+x_5+x_6+x_7+x_8+x_9+x_{11}+x_{12}+x_{14}+x_{15}) = \begin{bmatrix} 3 \\ \frac{28}{11} \end{bmatrix}$$

$u_{11}=1$, $u_{41}=1$, $u_{10,1}=1$, $u_{13,1}=1$

$u_{22}=1$, $u_{52}=1$, $u_{72}=1$, $u_{82}=1$, $u_{92}=1$, $u_{11,2}=1$, $u_{14,2}=1$, $u_{62}=1$, $u_{12,2}=1$, $u_{32}=1$, $u_{152}=1$

$$\underline{\theta}_1 = \frac{1}{4}(x_1+x_4+x_{10}+x_{13}) = \begin{bmatrix} 3 \\ \frac{1}{2} \end{bmatrix}, \underline{\theta}_2 = \frac{1}{11}(x_2+x_3+x_5+x_6+x_7+x_8+x_9+x_{11}+x_{12}+x_{14}+x_{15}) = \begin{bmatrix} 3 \\ \frac{28}{11} \end{bmatrix}$$

* چون $\underline{\theta}_i$ ها تغییری نکردند الگوریتم متوقف می‌شود.

پ)

Dunn index $D_m = \dfrac{d(C_1, C_2)}{\max\limits_{k=1,2} diam(C_k)}$

$d(C_1, C_2) = \min\limits_{x \in C_1, y \in C_2} = d(x_4, x_5) = 1$, $diam(C_1) = \max\limits_{x,y \in C_1} d(x,y) = d(x_1, x_{13}) = 6$

$diam(C_2) = \max\limits_{x \in C_1, y \in C_2} d(x,y) = d(x_2, x_{15}) = \sqrt{36+4} = \sqrt{40}$ $\longrightarrow D_m = \dfrac{1}{\sqrt{40}} \simeq 0.158$

Made with Goodnotes

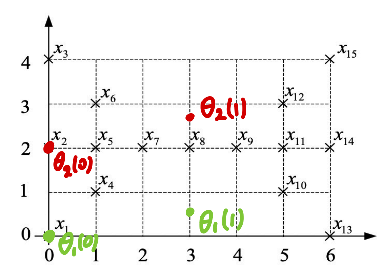
ادامه سوال 6 قسمت پ)

$$DB_2 = \frac{1}{2} \sum_{i=1}^{2} R_i \quad , \; s_i = diam(C_i) \rightsquigarrow$$ خودمون اینجوری تعریف می‌کنیم

$$R_1 = R_{12} = \frac{s_1 + s_2}{d_{12}} = \frac{6 + \sqrt{40}}{1} = 6 + \sqrt{40}$$

$$R_2 = R_{21} = R_{12}$$

$$\longrightarrow \boxed{DB_2 = R_1 = 6 + \sqrt{40}}$$



$$a_i = d_{avg}^{ps}(x_i, C_{c_i} - \{x_i\})$$

$$b_i = \min_{\substack{k=1,2 \\ k \neq c_i}} d_{avg}^{ps}(x_i, C_k)$$

$$a_1 = d_{avg}^{ps}(x_1, C_1 - \{x_1\}) = \frac{1}{3}\left(\sqrt{2} + \sqrt{26} + 6\right) \simeq$$

$$b_1 = d_{avg}^{ps}(x_1, C_2) = \frac{1}{11}\left(2 + \sqrt{5} + 2\sqrt{2} + \sqrt{13} + \sqrt{20} + \sqrt{29} + \sqrt{40} + \sqrt{10} + \sqrt{34} + 4 + \sqrt{52}\right)$$

همینجوری باید همه a_i ها و b_i ها رو حساب کنیم میشه:

$$s_i = \frac{b_i - a_i}{\max(b_i, a_i)} \; , \; -1 \leqslant s_i \leqslant 1$$

$$S_j = \frac{1}{n_j} \sum_{i : x_i \in C_j} s_i \longrightarrow S_m = \frac{1}{2} \sum_{j=1}^{2} S_j \; , \; S_m \in [-1, 1]$$

* خیلی طول می‌کشید حساب کردنش برای 15 داده به صورت دستی.

سؤال 7 -

الف)

$x_1=\begin{bmatrix}2\\3\end{bmatrix}$, $x_2=\begin{bmatrix}3\\5\end{bmatrix}$, $x_3=\begin{bmatrix}4\\7\end{bmatrix}$, $x_4=\begin{bmatrix}5\\9\end{bmatrix}$, $x_5=\begin{bmatrix}6\\11\end{bmatrix}$,

$t=2$
$m=2$
$N=5$
$L=2$

$\underline{\theta}_1(0)=\begin{bmatrix}3\\4\end{bmatrix}$, $\underline{\theta}_2(0)=\begin{bmatrix}5\\8\end{bmatrix}$

$$u_{ij}=\frac{1}{\sum \frac{d(x_i,\theta_j(t-1))}{d(x_i,\theta_k(t-1))}} \longrightarrow u_{i1}=\frac{1}{1+\frac{d(x_i,\theta_1(t-1))}{d(x_i,\theta_2(t-1))}}$$

به همین ترتیب u_{ij} ها را محاسبه می کنیم:

$u_{11}=0.8$
$u_{21}=0.78$
$u_{31}=0.31$
$u_{41}=0.16$
$u_{51}=0.29$

$u_{12}=0.19$
$u_{22}=0.21$
$u_{32}=0.69$
$u_{42}=0.84$
$u_{52}=0.7$

$$\underline{\theta}_j=\frac{\sum_{i=1}^{5}u_{ij}x_i}{\sum_{i=1}^{5}u_{ij}} \longrightarrow \underline{\theta}_1(1)=\begin{bmatrix}3.29\\5.59\end{bmatrix} \quad \underline{\theta}_2(1)=\begin{bmatrix}4.63\\8.25\end{bmatrix}$$

ب)

$$PC=\frac{1}{5}\sum_{i=1}^{5}\sum_{j=1}^{2}u_{ij}^2=\frac{1}{5}\times 3.206=\boxed{0.6412}$$

$$PE=-\frac{1}{5}\sum_{i=1}^{5}\sum_{j=1}^{2}u_{ij}\log_a u_{ij}$$

* a را 2 در نظر می گیریم.

$$\longrightarrow PE=-\frac{1}{5}\times -3.871=\boxed{0.774}$$