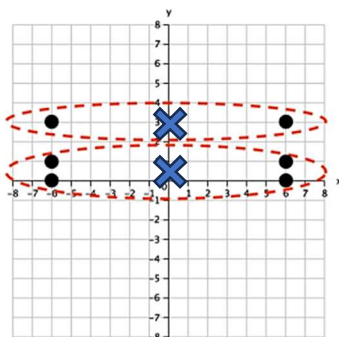
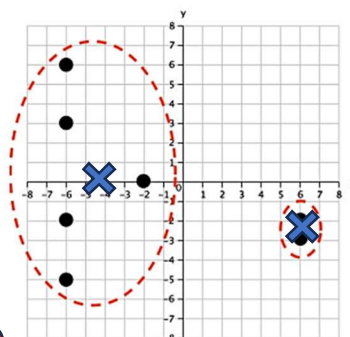
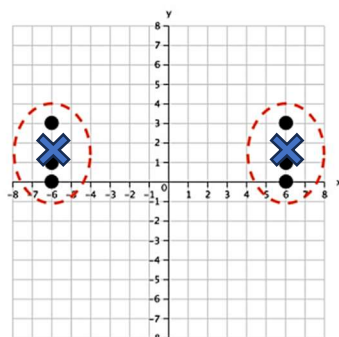
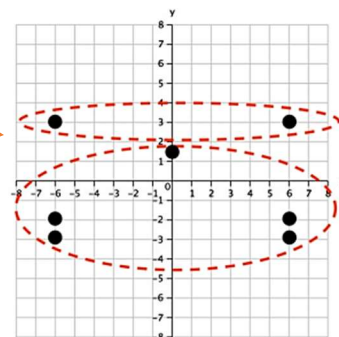


1. ما می‌خواهیم نقاط دو بعدی را به دو خوشه تقسیم کنیم، بنابراین الگوریتم لُویِد برای خوشه‌بندی k-means را تا همگرایی اجرا می‌کنیم. کدام یک از خوشه‌های زیر می‌تواند تولید شود؟ (نقاط داخل یک بیضی به یک خوشه تعلق دارند).

مراکز اولیه  
ای که می  
توانند به این  
خوشه بندی ها  
منجر شوند  
برای هر شکل  
رسم شده اند

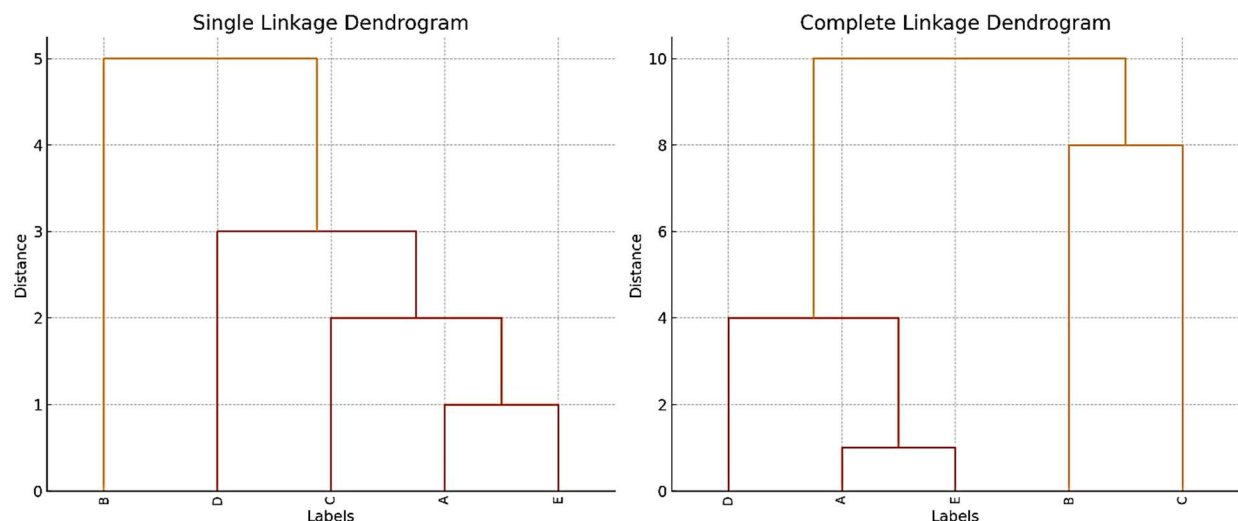


چرا این خوشه بندی نمی  
تواند هرگز اتفاق بیفتد؟  
- مرکز های دو خوشه را  
حساب کنید، بلافاصله  
assignment نقاط در  
گام بعدی تغییر میکند و  
خوشه بندی به این شکل  
همگرا نمی شود



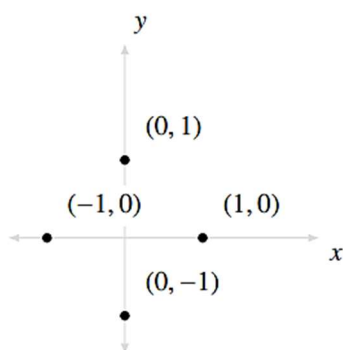
distance	A	B	C	D	E
A	0	9	2	3	1
B		0	8	6	5
C			0	7	10
D				0	4
E					0

2. یک مجموعه داده شامل A, B, C, D, E با ماتریس فاصله زیر داده شده است. الگوریتم های complete-link و average-link hierarchical clustering به مجموعه داده کنید. نمودارهای dendrogram را رسم کنید.



3. الگوریتم خوشه‌بندی سلسله‌مراتبی را روی مجموعه‌ای از چهار نقطه در نظر بگیرید، به‌طوری‌که در صورت وجود تساوی،

آن را به‌صورت دلخواه بشکنیم. اگر الگوریتم را متوقف کنیم زمانی که تنها دو خوشه باقی مانده‌اند، کدام یک از روش‌های لینک‌دهی زیر تضمین می‌کند که خوشه‌های حاصل متوازن باشند (هر کدام دارای دو نقطه نمونه باشند)؟ تمام موارد صحیح را انتخاب کنید.



A: Complete linkage, B: Single linkage, C: Centroid linkage

D: Average linkage

همه موارد منجر به خوشه بندی متوازن می شوند به جز single linkage.

تحلیل single linkage: فواصل نقاطی که در هر ربع محور مختصات قرار دارند یکسان است. فرض کنید اول به طور تصادفی  $(0, 1)$  و  $(1, 0)$  انتخاب شده باشند و با هم ادغام شوند. چون فاصله  $(-1, 0)$  تا خوشه  $\{(0, 1), (1, 0)\}$  با فاصله  $(0, -1)$  تا  $(-1, 0)$  و همچنین فاصله  $(0, -1)$  تا خوشه  $\{(0, 1), (1, 0)\}$  یکسان است، ممکن است  $(-1, 0)$  با خوشه فعلی  $\{(0, 1), (1, 0)\}$  ادغام شود که منجر به یک خوشه سه تایی و یک خوشه تک عنصری میشود.

تحلیل complete linkage: فرض کنید به صورت تصادفی  $(0, 1)$  و  $(1, 0)$  انتخاب شده باشند. در این حالت دیگر  $(-1, 0)$  فاصله بیشتری با خوشه دو عنصری  $\{(1, 0), (0, 1)\}$  دارد تا داده  $(0, -1)$  و بنابراین با نقطه  $(0, -1)$  ادغام می شود. همین حالت برای  $(0, -1)$  نیز برقرار است. پس complete منجر به دو خوشه متوازن می شود.

تحلیل average linkage: مشابه با complete (میانگین فاصله  $(-1,0)$  تا خوشه  $\{(0,1), (1,0)\}$  بیشتر از فاصله تا  $(0,-1)$  است).

تحلیل centroid: مشابه با average و complete.

TID	Date	items_bought
T100	10/15/04	{K, A, D, B}
T200	10/15/04	{D, A, C, E, B}
T300	10/19/04	{C, A, B, E}
T400	10/22/04	{B, A, D}

4. یک پایگاه داده شامل ۴ تراکنش است که در زیر نشان داده شده‌اند.

با فرض حداقل سطح پشتیبانی  $\text{min\_support} = 60\%$  و حداقل سطح اطمینان  $\text{min\_confidence} = 80\%$ :

(الف) تمام مجموعه‌های آیتم‌های پرتکرار را با استفاده از الگوریتم

Apriori پیدا کنید. کار خود را نشان دهید—فقط نشان دادن پاسخ نهایی قابل قبول نیست. برای هر تکرار، مجموعه‌های آیتم‌های کاندید و مجموعه‌های آیتم‌های پرتکرار قابل قبول را نمایش دهید.

item	support
A	4/4
B	4/4
C	2/4
D	3/4
E	2/4
K	1/4

item	support
A,B	4/4
A,D	3/4
B,D	3/4

item	support
A,B,D	3/4

(ب) تمام قوانین ارتباط قوی را همراه با مقادیر اطمینان (confidence) آن‌ها فهرست کنید.

Confidence (A  $\rightarrow$  B):  $\text{support}(A,B)/\text{support}(A)=1$ , Confidence (B  $\rightarrow$  A):  $\text{support}(A,B)/\text{support}(B)=1$

Confidence (A  $\rightarrow$  D):  $1/1=1$ , confidence (D  $\rightarrow$  A):  $(3/4)/(3/4)=1$

Confidence (A  $\rightarrow$  B,D):  $(3/4)/(1)=0.75$ , confidence (B  $\rightarrow$  A,D):  $(3/4)/(1)=3/4$ , confidence (D  $\rightarrow$  A,B)=1

$\text{Confidence}(A, B \rightarrow D) = 3/4$ ,  $\text{Confidence}(B, D \rightarrow A) = 1$ ,  $\text{Confidence}(A, D \rightarrow B) = 1$

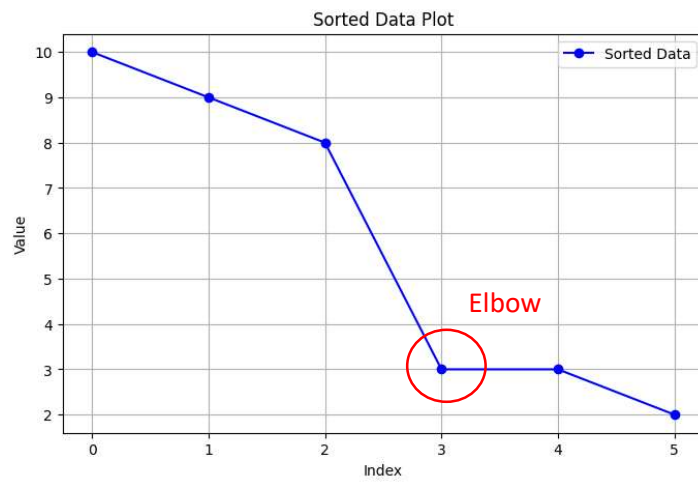
5. یک دیتاست متشکل از اشیاء A، B، C، D، E و F با ماتریس فاصله زیر داده شده است:

distance	A	B	C	D	E	F
A	0	1	2	4	6	7
B		0	3	8	9	10
C			0	11	12	13
D				0	14	15
E					0	16
F						0

با استفاده از الگوریتم DBSCAN و روش k-distance plot، بهترین مقدار  $\epsilon$  را برای  $\text{minpoints}=3$  پیدا کنید. برای این مقدار  $\epsilon$  خوشه بندی نهایی را محاسبه کنید.

Set  $k=2$

distance	A	B	C	D	E	F	k-distance
A	0	1	2	4	6	7	$\text{Max}(1,2)=2$
B		0	3	8	9	10	$\text{Max}(3,1)=3$
C			0	11	12	13	$\text{Max}(2,3)=3$
D				0	14	15	$\text{Max}(4,8)=8$
E					0	16	$\text{Max}(6,9)=9$
F						0	$\text{Max}(7,10)=10$



Set  $\text{eps} = 3$