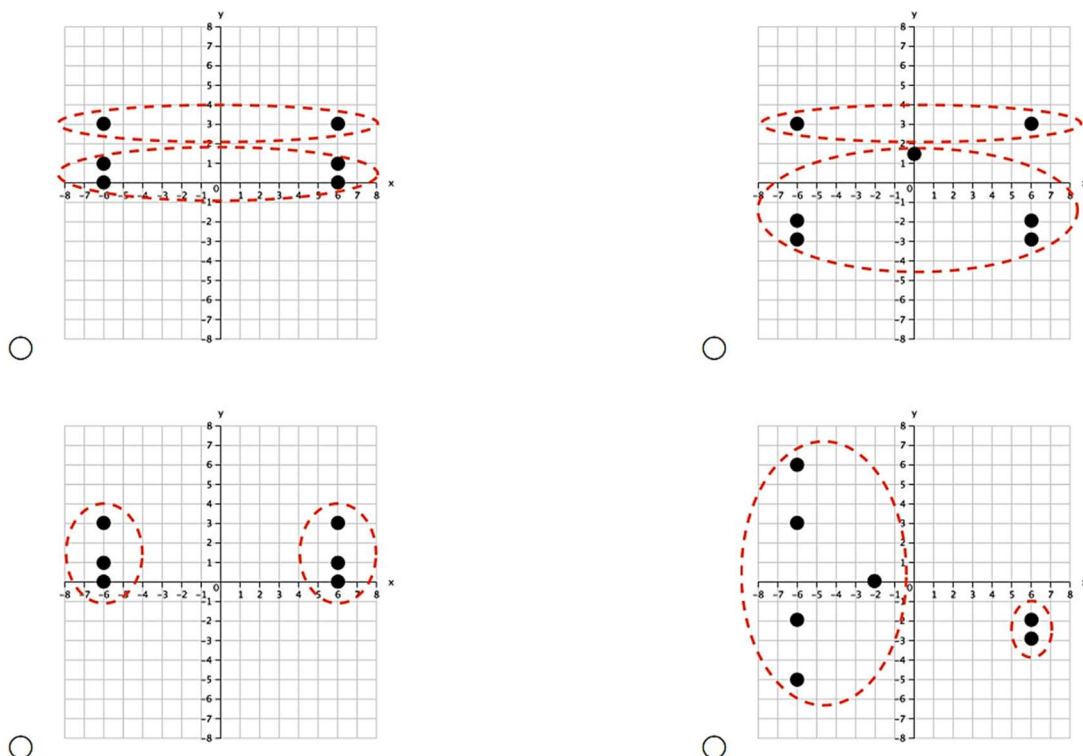


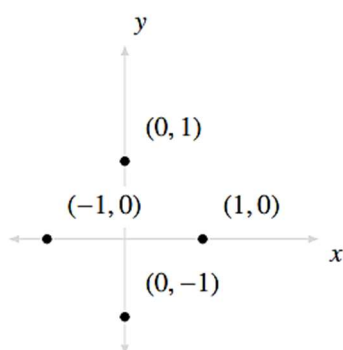
1. ما می‌خواهیم نقاط دو بعدی را به دو خوشه تقسیم کنیم، بنابراین الگوریتم لُویِد برای خوشه‌بندی k-means را تا همگرایی اجرا می‌کنیم. کدام یک از خوشه‌های زیر می‌تواند تولید شود؟ (نقاط داخل یک بیضی به یک خوشه تعلق دارند).



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

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

3. الگوریتم خوشه‌بندی سلسله‌مراتبی را روی مجموعه‌ای از چهار نقطه در نظر بگیرید، به‌طوری‌که در صورت وجود تساوی، آن را به صورت دلخواه بشکنیم. اگر الگوریتم را متوقف کنیم زمانی که تنها دو خوشه باقی مانده‌اند، کدام یک از روش‌های لینک‌دهی زیر تضمین می‌کنند 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

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 آیت‌های 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 آیت‌های کاندید و مجموعه‌های آیت‌های پرتکرار قابل قبول را نمایش دهید.

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

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 ϵ را برای $\text{minpoints}=3$ پیدا کنید. برای این مقدار ϵ خوشه بندی نهایی را محاسبه کنید.