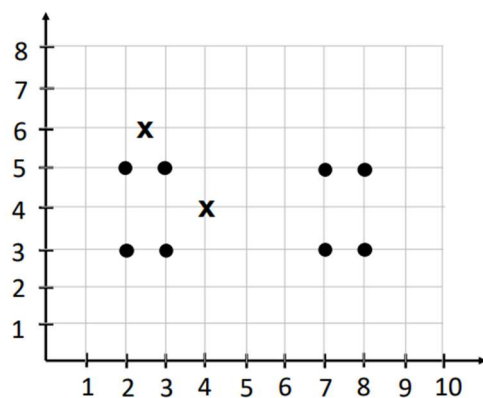


1. یک مجموعه داده شامل اشیاء A, B, C, D, E, F با ماتریس فاصله زیر داده شده است:

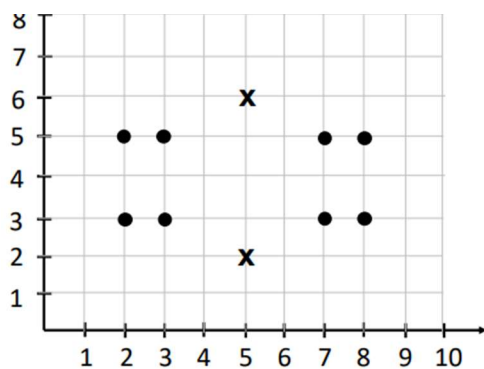
distance	A	B	C	D	E	F
A	0	1	2	4	6	7
B		0	3	8	9	10
C			0	11	12	13
D				0	14	15
E					0	16
F						0

فرض کنید الگوریتم DBSCAN برای این مجموعه داده با $\text{MINPOINTS} = 3$ و $\text{epsilon} = \varepsilon = 5$ اجرا شود. الگوریتم DBSCAN چند خوشه باز می‌گرداند و چه داده‌هایی در هر خوشه هستند؟ کدام اشیاء در نتیجه خوشه‌بندی قبلی به عنوان نقاط برون‌زای (outliers) و نقاط مرزی (borderpoints) شناخته می‌شوند؟



Initial state

شکل ۲. مقداردهی اولیه قسمت الف

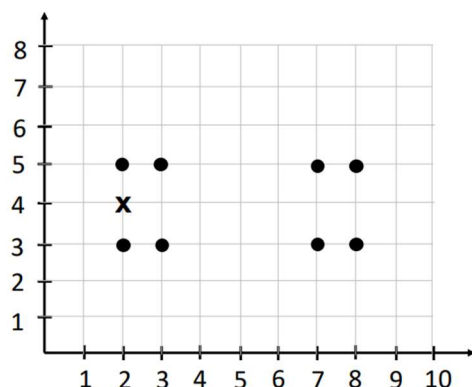


شکل ۱. مقداردهی قسمت ج

2. فرض کنید تعداد خوشه‌ها $k=2$ برای تمام قسمت‌های زیر باشد.

(الف) مراحل الگوریتم k-means را گام به گام طی کنید، از مقداردهی اولیه‌ای که در نمودار سمت چپ بالا در جعبه زیر نشان داده شده است شروع کنید. نقاط، داده‌های مشاهده شده را نشان می‌دهند. با دو علامت 'x' مکان مراکز خوشه‌ها و با رسم دایره داده‌های هر خوشه را در هر تکرار الگوریتم k-means مشخص کنید. مراکز خوشه‌های اولیه در حالت اولیه از پیش نشان داده شده‌اند. (ب) مقدار عددی inertia برای خوشه‌بندی به دست آمده در قسمت (الف) پس از اتمام اجرای الگوریتم چیست؟

(ج) مانند قسمت (الف)، مراحل الگوریتم k-means را گام به گام طی کنید، از مقداردهی اولیه‌ای که در نمودار سمت چپ بالا نشان داده شده است شروع کنید. مقدار inertia در این حالت چند است؟ کدام مقداردهی اولیه بهتر است؟



د) نقاط داده‌ای را که با دایره‌های سیاه در نمودار زیر نشان داده شده‌اند در نظر بگیرید. ما یک مرکز خوشه را با علامت X در مختصات (2,4) رسم کرده‌ایم. مرکز خوشه دوم را به گونه‌ای رسم کنید که شرط زیر را برآورده کند:

وقتی مراکز خوشه‌ها را در دو نقطه مقداردهی اولیه کنیم و الگوریتم k-means را تا همگرایی اجرا کنیم، وضعیت نهایی به گونه‌ای خواهد بود که یک خوشه شامل تمام نقاط داده باشد و خوشه دیگر هیچ نقطه‌ای نداشته باشد.

۳. کدام یک از روش‌های خوشه بندی داده شده داده‌های موجود شکل زیر را به دو خوشه دایره قرمز و خط افقی آبی تقسیم می‌کند؟ هر نقطه در دایره و خط یک نقطه داده است. در تمام گزینه‌هایی که شامل خوشه‌بندی سلسله‌مراتبی هستند، الگوریتم تا زمانی اجرا می‌شود که دو خوشه به دست آید.



- الف) خوشه بندی سلسه مراتبی complete-link
- ب) خوشه بندی سلسه مراتبی single-link
- ج) خوشه بندی kmeans

۴. با استفاده از روش Fp-growth آیتم های پرتکرار با min-support count=3 را بدست آورید:

TID	Items Bought
1	a, b, f
2	b, g, c, d
3	h, a, c, d, e
4	a, d, p, e
5	a, b, c
6	a, b, q, c, d
7	a
8	a, m, b, c
9	a, b, n, d
10	b, c, e, m