

1

Class A $\rightarrow \{p1,p2,p3\}$

Precision(i,j) class i cluster j: $= \frac{N_{ij}}{N_j}$

Class B $\rightarrow \{p4,p5,p6,p7,p8\}$

Recall(i,j) class i cluster j: $= \frac{N_{ij}}{N_i}$

$$F(i,j) = \frac{2R(i,j) * P(i,j)}{R(i,j) + P(i,j)}$$

cluster #1 {p1,p2,p3,p4,p5,p6,p7,p8}

Class A

$$P(A,1) = \frac{N_{A1}}{N_1} = \frac{3}{8} = 0.375, R(A,1) = \frac{N_{A1}}{N_A} = \frac{3}{3} = 1, F(A,1) = \frac{2 * 0.375}{1.375} = 0.545$$

Class B

$$P(B,1) = \frac{N_{B1}}{N_1} = \frac{5}{8} = 0.625, R(B,1) = \frac{N_{B1}}{N_B} = \frac{5}{5} = 1, F(B,1) = \frac{2 * 0.625}{1.625} = 0.769$$

cluster #2 {p1,p2,p4,p5}

Class A

$$P(A,2) = \frac{N_{A2}}{N_2} = \frac{2}{4} = 0.5, R(A,2) = \frac{N_{A2}}{N_A} = \frac{2}{3} = 0.66$$

$$F(A,2) = \frac{2 * 0.66 * 0.5}{0.66 + 0.5} = 0.568$$

Class B

$$P(B,2) = \frac{N_{B2}}{N_2} = \frac{2}{4} = 0.5, R(B,2) = \frac{N_{B2}}{N_B} = \frac{2}{5} = 0.4$$

$$F(B,2) = \frac{2 * 0.5 * 0.4}{0.5 + 0.4} = 0.444$$

cluster #3 {p3,p6,p7,p8}

Class A

$$P(A,3) = \frac{N_{A3}}{N_3} = \frac{1}{4} = 0.25, R(A,3) = \frac{N_{A3}}{N_A} = \frac{1}{3} = 0.33$$

$$F(A,3) = \frac{2 * 0.25 * 0.33}{0.25 + 0.33} = 0.284$$

Class B

$$P(B,3) = \frac{N_{B3}}{N_3} = \frac{3}{4} = 0.75, R(B,3) = \frac{N_{B3}}{N_B} = \frac{3}{5} = 0.6$$

$$F(B, 3) = \frac{2 * 0.75 * 0.6}{0.75 + 0.6} = 0.666$$

cluster #4 {p1,p2}

Class A

$$P(A, 4) = 1, R(A, 4) = 0.66, F(A, 4) = 0.795$$

Class B

$$P(B, 4) = 0, R(B, 4) = 0, F(B, 4) = 0$$

cluster #5 {p4,p5}

Class A

$$P(A, 5) = 0, R(A, 5) = 0, F(A, 5) = 0$$

Class B

$$P(B, 5) = 1, R(B, 5) = 0.4, F(B, 5) = 0.571$$

cluster #6 {p3,p6}

Class A

$$P(A, 6) = 0.5, R(A, 6) = 0.33, F(A, 6) = 0.397$$

Class B

$$P(B, 6) = 0.5, R(B, 6) = 0.2, F(B, 6) = 0.285$$

cluster #7 {p7,p8}

Class A

$$P(A, 7) = 0.0, R(A, 7) = 0.0, F(A, 7) = 0$$

Class B

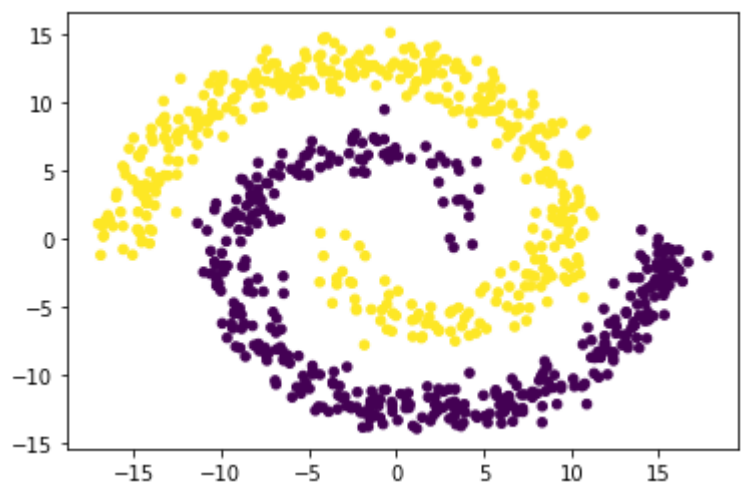
$$P(B, 7) = 1, R(B, 7) = 0.4, F(B, 7) = 0.571$$

2

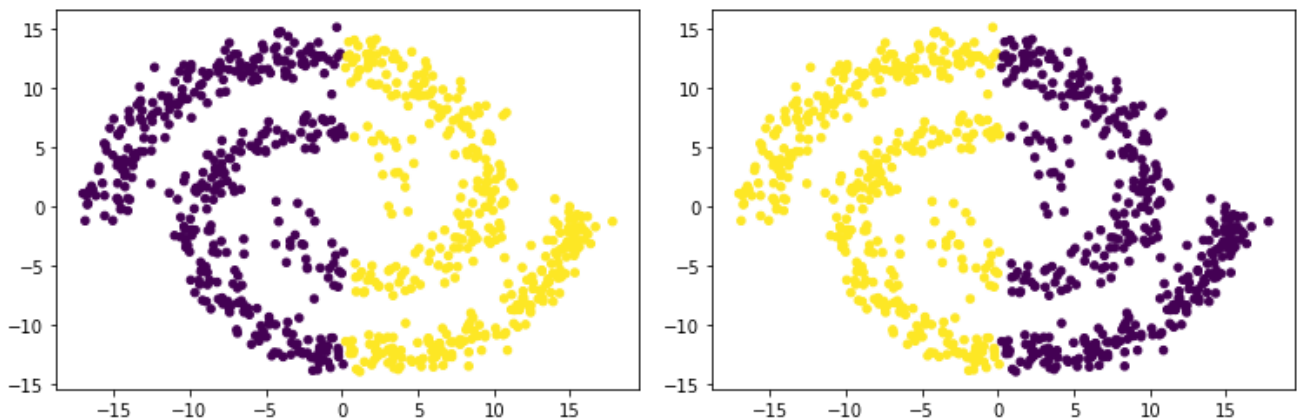
3

به دلیل اینکه هدف KMeans، مینیمایز کردن مقدار SSE است، میتواند با داده هایی با شکل های غیرعادی به مشکل برخورد کند، برای مثال من از داده spiral استفاده میکنم (کد مربوط به تولید و نمایش این داده و نمودار ها در فایل 3.ipynb قرار دارد)

نمونه این داده:



نمونه اجراهای متناوب بر روی این داده و نتیجه مشابه در هر اجرا



قدرت: به دلیل نحوه محاسبه مرکز کلاستر (میانگین) سرعت نسبتاً بیشتری نسبت به k-medoids که از میانه استفاده می‌کند دارد

ضعف: مقاومت کمتر در مقابل outliers

به دلیل اینکه مقادیر خیلی متفاوت تاثیر کمتری بر میانه نسبت به میانگین میگذارد، k-medoids مقاومت بیشتری نسبت به outliers دارد

ضعف: نیاز اولیه به تعداد کلاسترها دارند، در صورتی که روش سلسله مراتبی اینگونه نیست و میتوان تعداد مناسب کلاستر را با تفسیر و بررسی یک خوشه‌بندی سلسله مراتبی به دست آورد.

قدرت: از لحاظ محاسباتی (سرعت و فضا) سبک‌تر از روش‌های سلسله مراتبی هستند.

همچنین روش‌های سلسله مراتبی درمقابل شکل‌های hyper-spherical (مانند دایره در فضای دو بعدی یا کره در فضای سه بعدی) ضعیف‌تر عمل می‌کنند.