به نام خدا



دانشگاه صنعتی امیرکبیر

(پلی‌تکنیک تهران)

درس شناسایی آماری الگو

استاد رحمتی

تمرین چهارم

علیرضا مازوچی

۴۰۰۱۳۱۰۷۵

سوال ۱

a و b و c) با درنظرگرفتن مرکزهای اولیه، داده‌های هر خوشه عبارت است از:

بر همین مبنا مرکز جدید سه خوشه تغییر خواهد کرد:

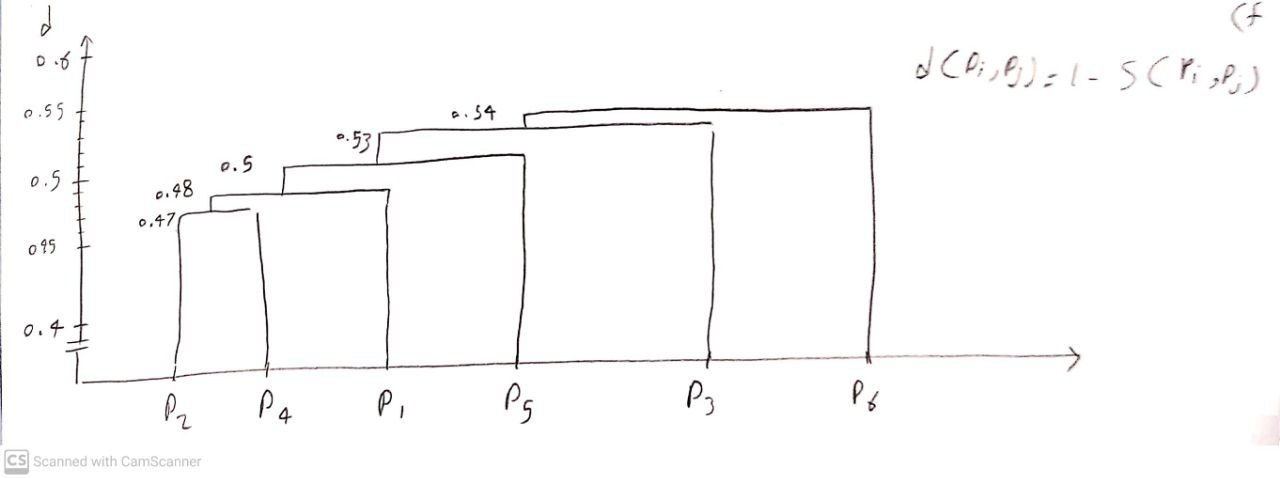
d) باتوجه به مراکز جدید خوشه‌ها، داده‌های خوشه‌ها در دومین گام عبارت است از:

باتوجه به اینکه جای داده‌ها تغییر نکرده است پس مراکز خوشه‌ها هم تغییری نخواهد کرد و لذا الگوریتم در پایان گام دوم خاتمه می‌یابد.

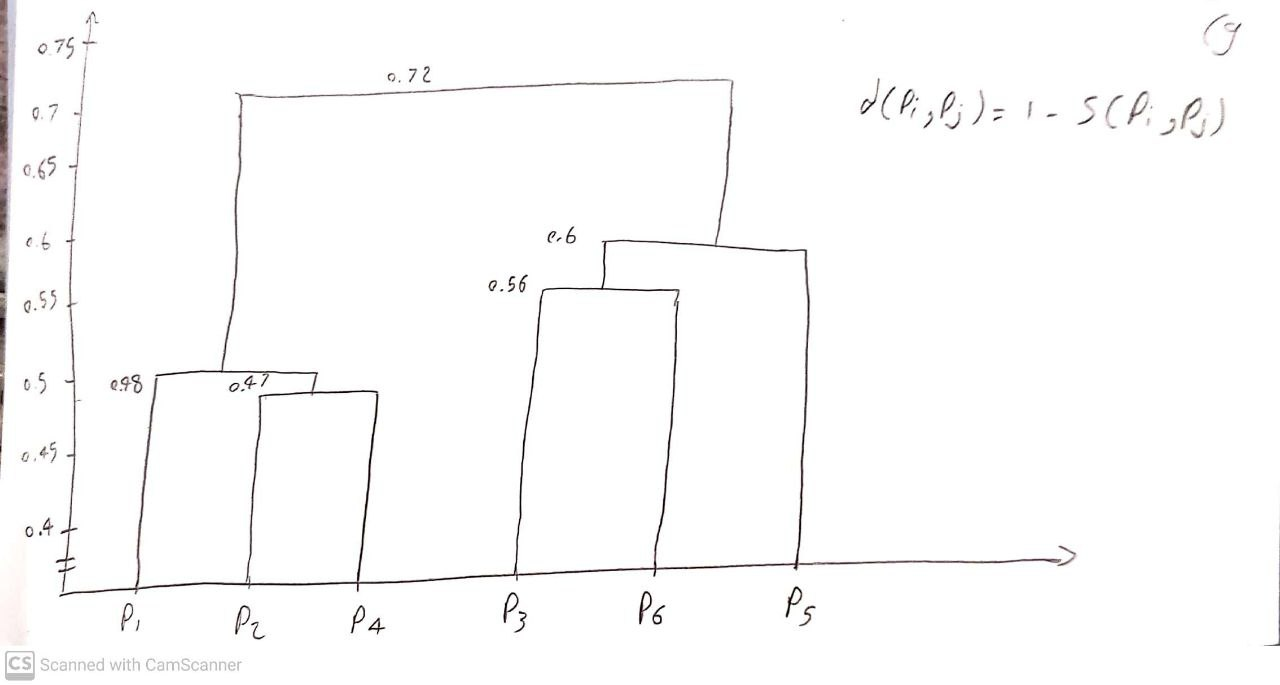
e) حالت بهینه سراسری زمانی رخ می‌دهد که یک خوشه دارای داده‌ی ۳- و ۱ باشد و خوشه دیگر دارای داده ۸. اگر شرایط به گونه‌ای پیش برود که یک خوشه بدون داده بماند و یا اینکه یک خوشه دارای ۸ و ۱ و دیگری دارای ۳- باشد به بهینه سراسری دست پیدا نخواهیم کرد. لذا می‌توان این شرایط را در نظر گرفت:

با اشتراک‌گیری از شرایط بالا به بازه‌های زیر می‌رسیم:

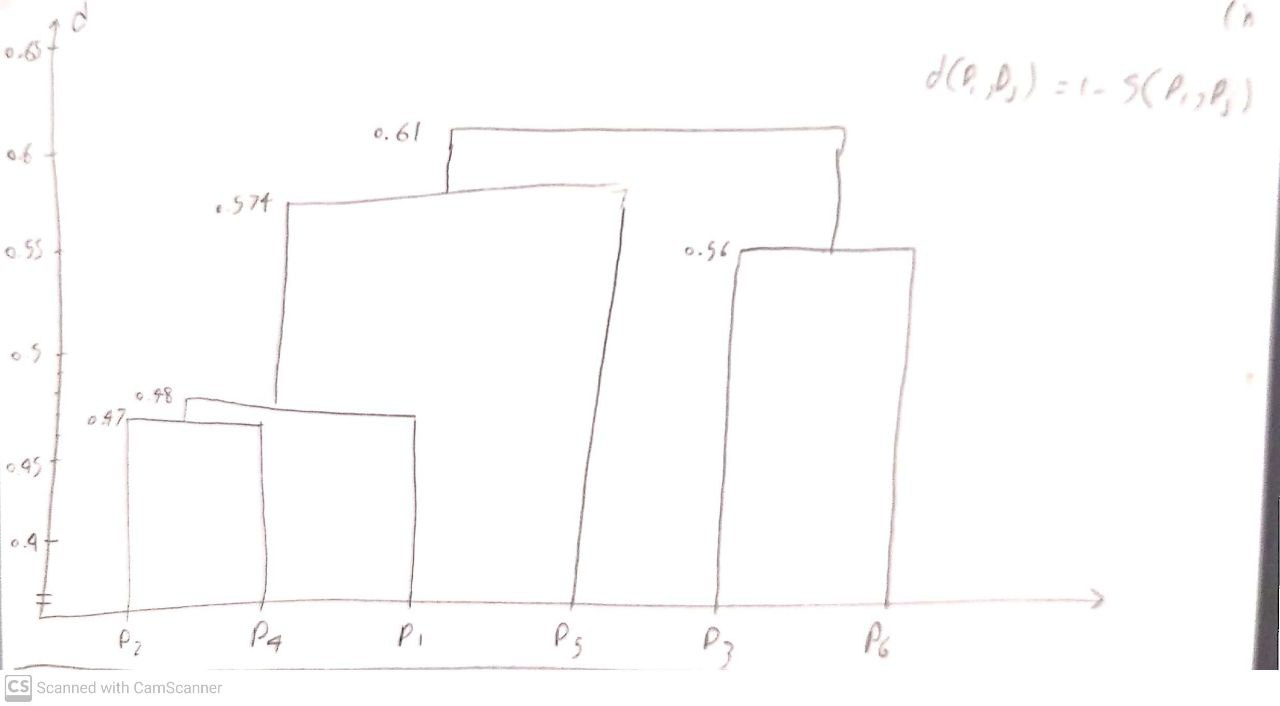
f)



g)



h)



سوال ۲

b)

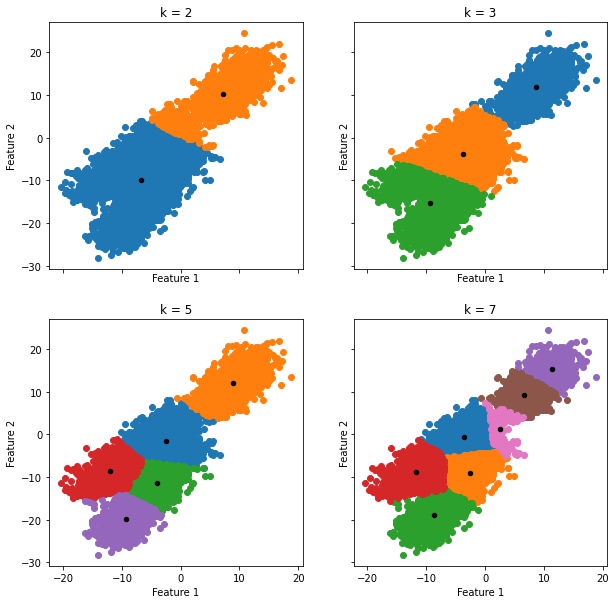
c)

d) به صورت عادی این مراکز، مراکزی هستند که همگرا شده‌اند و اگر الگوریتم خوشه‌بندی با این مراکز شروع شود حد آستانه همگرایی شکسته می‌شود و در همان گام اول الگوریتم به پایان می‌رسد. یک احتمال دیگر هم که وجود دارد آن است که تعداد گام درنظر گرفته شده برای خوشه‌بندی اول کافی نباشد. در این صورت ممکن است در دفعه دوم خوشه‌بندی مراکز کمی تغییر کند.

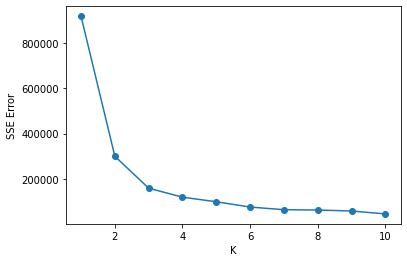
e) اگر داده‌های بابرچسب داشته باشیم، استفاده از یک الگوریتم نظارت‌شده معمولا به نتایج بهتری منجر می‌شود؛ چراکه در این حالت دانش بیشتری وجود دارد که یک الگوریتم نظارت‌شده از آن استفاده می‌کند ولی یک الگوریتم بدون نظارت آن را نادیده می‌گیرد. در مقایسه یک الگوریتم غیرنظارت‌شده دیگر با k-means هم به قطعیت نمی‌توان نظری داد و بسته به نوع داده‌ها و پارامتر‌های هر الگوریتم ممکن است k-means بهتر باشد و ممکن است الگوریتم دیگر عملکرد بهتری داشته باشد.

سوال ۳

a)



b)



در این نمودار انتظار می‌رود که با افزایش مقدار k خطای خوشه‌بندی هم کاهش یابد اما مقدار کاهش روند ثابتی را ندارد؛ در ابتدا کاهش شدید است تا به نقطه‌ای برسیم که تعداد خوشه‌های متناسب با داده‌ها بدست آید و از آنجا به بعد کاهش کمتری رخ خواهد داد. لذا یک شکستگی در نمودار دیده می‌شود که تعداد خوشه مناسب را نمایان می‌کند. در این نمودار هم ۳ خوشه به نظر مناسب می‌آیند.

سوال ۴

سوال ۵