



دانشگاه صنعتی امیر کبیر
(پلی تکنیک تهران)

گزارش کد تمرین سوم

درس هوش محاسباتی

استاد عبادزاده

عطیه براتی نیا ۹۶۳۱۰۱۰

سوال ۱

الگوریتم بدین صورت کار میکند که ابتدا تعداد خوشه‌ها را تعیین میکنیم (C) و سپس مراکز خوشه را مقداردهی میکنیم، مقداردهی اولیه در این نمونه بدین صورت انجام شده که نقاط $1/(i+1)$ (i از یک تا C)، به عنوان مراکز خوشه تعیین شده اند (بدین دلیل این گونه مقداردهی شده تا مراکز با هم متفاوت باشند). و از اینجا به بعد قدم های زیر تکرار میشود:

۱. مقدار تعلق هر داده به هر خوشه را به دست می‌آوریم.
۲. مراکز خوشه را بر طبق تعلق‌های جدید مجدد به دست می‌آوریم.
۳. آنترپی را محاسبه میکنیم.
۴. اگر آنترپی نسبت به مرحله پیش کمتر از ۰.۰۰۵ تغییر داشته، الگوریتم را تمام کن در غیر اینصورت به ۱ برو.

الگوریتم بالا در حالتی است که تعداد خوشه‌ها را داشته باشیم، از آنجایی که ممکن است تعداد خوشه ها را نداشته باشیم از $C=1$ الگوریتم بالا را شروع کرده و آنترپی انتهایی را نگه میداریم، سپس یکی یکی به تعداد خوشه می‌افزاییم تا جایی که اضافه کردن خوشه تغییر محسوسی در آنترپی نداشته باشد. آن وقت تعداد خوشه را نیز یافته‌ایم، در این مساله تعداد خوشه تا زمانی اضافه میشود که تغییر آنترپی از ۵ بیشتر باشد. فرمول های محاسبه‌ی مقدار تعلق و مراکز خوشه از لینکی که در صورت سوال آمده، استفاده شده است.

سوال ۲

برای تابع هزینه از معیار آنترپی استفاده کرده‌ام، این معیار مشخص میکند که داده‌ها چقدر از صفر و یک دورند، هر چه داده‌ها به صفر و یک نزدیک باشند یعنی خوشه بندی صحیح را پیدا کرده‌ایم و هرچه از مقدار صفر و یک دورتر باشند و به ۰.۵ نزدیک تر باشند یعنی داده‌ها هنوز به خوشه‌ی خاصی تعلق پیدا نکرده‌اند.

$$\text{معیار آنترپی} = \frac{-\sum_{i=1}^C \sum_{k=1}^N u_{ik} \ln u_{ik}}{C}$$

با افزایش تعداد خوشه‌ها معیار آنترپی کاهش میابد چون تعداد مراکز افزایش میابد و هر داده به خوشه نزدیک تر تعلق پیدا میکند و در نتیجه بی‌نظمی کاهش میابد و آنترپی کم میشود.

سوال ۳

مقدار m باید بزرگتر از ۱ باشد تا بتوان از رابطه E مشتق گرفت و روابط مرکز و تعلق را به دست آورد که در اینجا این مقدار ۱.۵ و ۲ تست شدند که ۱.۵ جواب بهتری به ما داد و این مقدار برای m تعیین شد.

سوال ۴

یکی از معیارهای خوشه بندی خوب کاهش معیار بازسازی یا کاهش معیار آنتروپی است، هرچند که با بیشتر شدن تعداد خوشه‌ها آنتروپی کاهش میابد ولی هزینه محاسباتی افزایش میابد. هدف پیدا کردن تعداد خوشه کمترین و بهینه‌ترین است. به همین دلیل یکی از روش‌ها روش *elbow* است که در این روش جایی که آنتروپی بعد از آن کاهش چشمگیری ندارد را به عنوان تعداد خوشه بهینه در نظر میگیرد.

پس از اجرای الگوریتم FCM

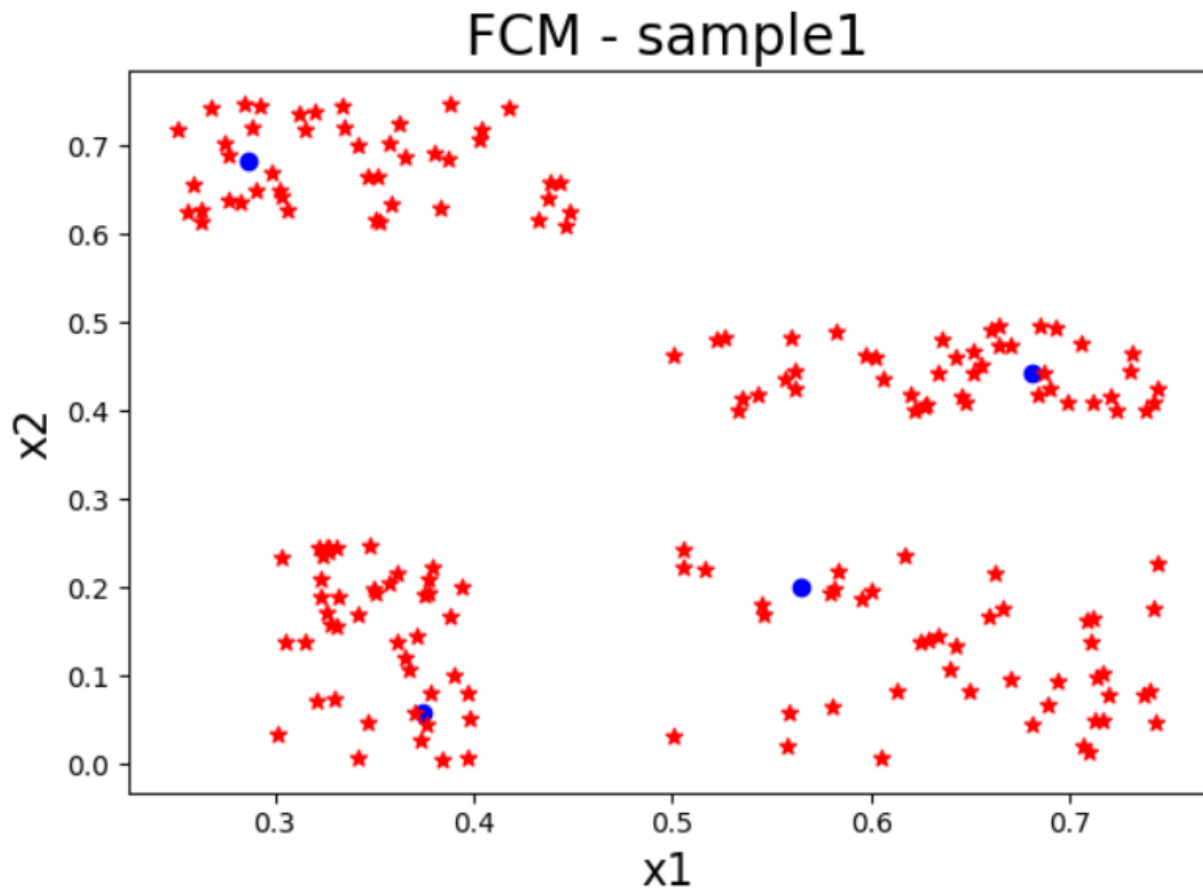
الگوریتم به صورتی نوشته شده است که ما تعداد خوشه‌بندی را نداریم بنابراین از $C=1$ شروع کرده و آنتروپی انتهایی هر خوشه‌بندی را نگه می‌داریم، سپس یکی یکی به تعداد خوشه می‌افزاییم تا جایی که اضافه کردن خوشه تغییر محسوسی در آنتروپی نداشته باشد. آن وقت تعداد خوشه را یافته‌ایم، در این مساله تعداد خوشه تا زمانی که تغییر آنتروپی از ۵ بیشتر است زیاد میشود.

الگوریتم بدین صورت کار میکند که ابتدا تعداد خوشه‌ها را تعیین میکنیم (C) و سپس مراکز خوشه را مقداردهی میکنیم، مقداردهی اولیه در این نمونه بدین صورت انجام شده که نقاط $1/(i+1)$ (i از یک تا C)، به عنوان مراکز خوشه تعیین شده اند (بدین دلیل این گونه مقداردهی شده تا مراکز با هم متفاوت باشند.) و از اینجا به بعد قدم های زیر تکرار میشود:

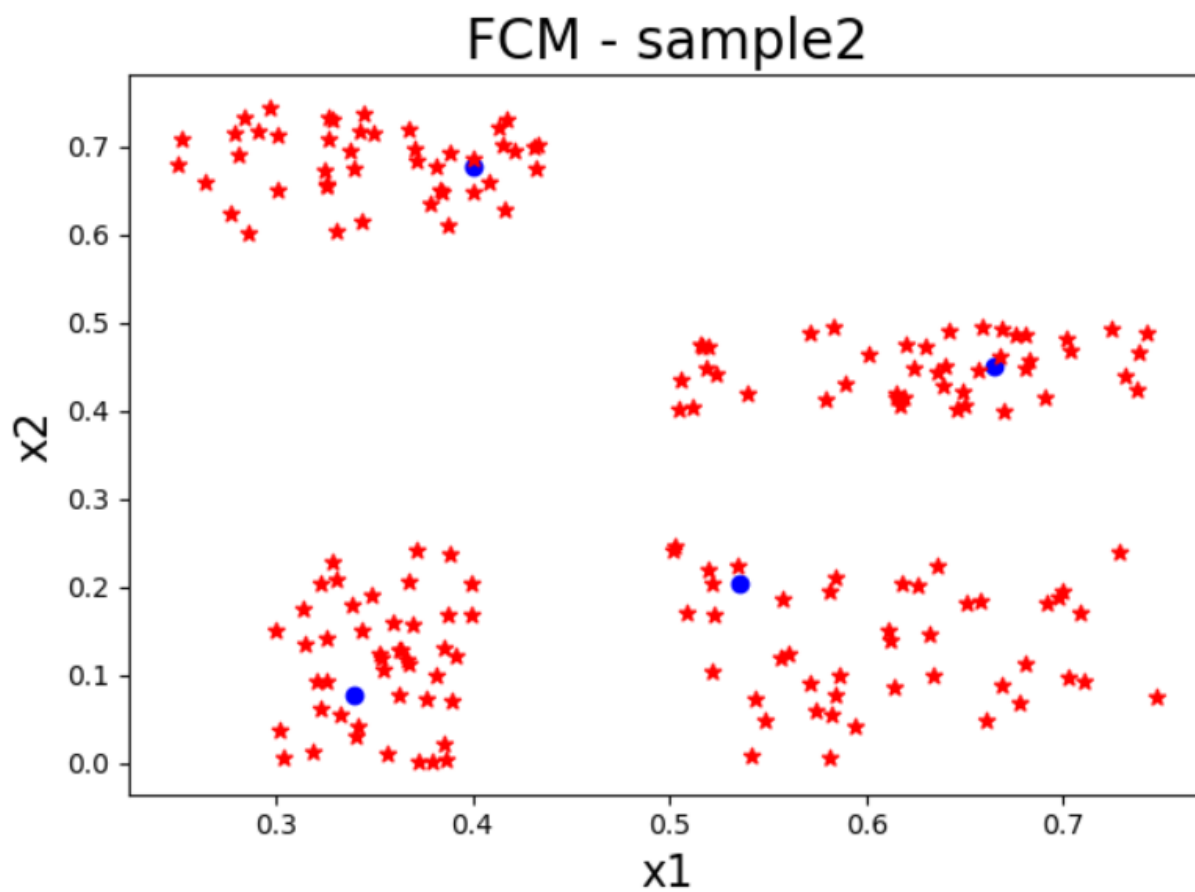
۱. مقدار تعلق هر داده به هر خوشه را به دست می‌آوریم.
۲. مراکز خوشه را بر طبق تعلق‌های جدید مجدد به دست می‌آوریم.
۳. آنتروپی را محاسبه میکنیم.
۴. اگر آنتروپی نسبت به مرحله پیش کمتر از ۰.۰۰۵ تغییر داشته، الگوریتم را تمام کن در غیر اینصورت به ۱ برو.

این مقادیر آنتروپی بعد از سعی و خطای بسیار به دست آمد.

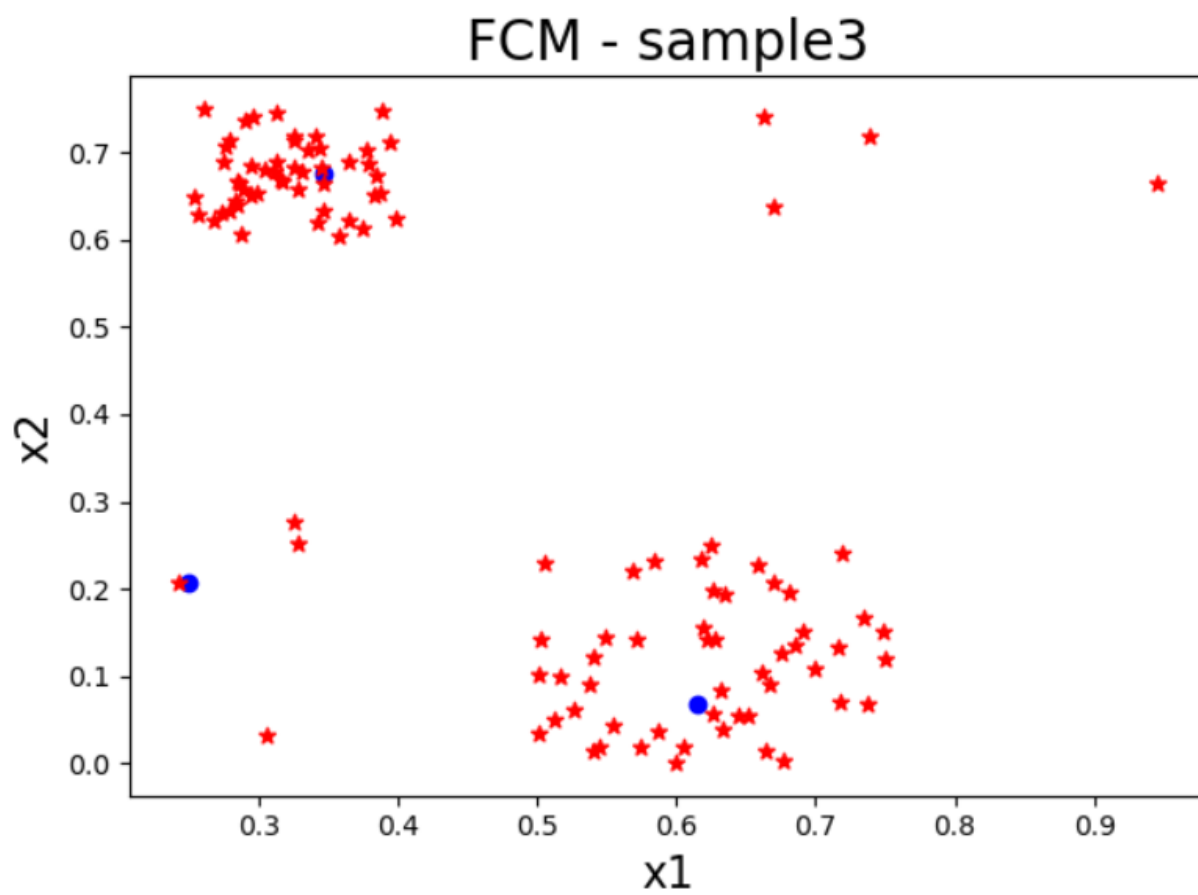
*تعداد خوشه‌ها با الگوریتم و طبق آنتروپی به دست آمده است.



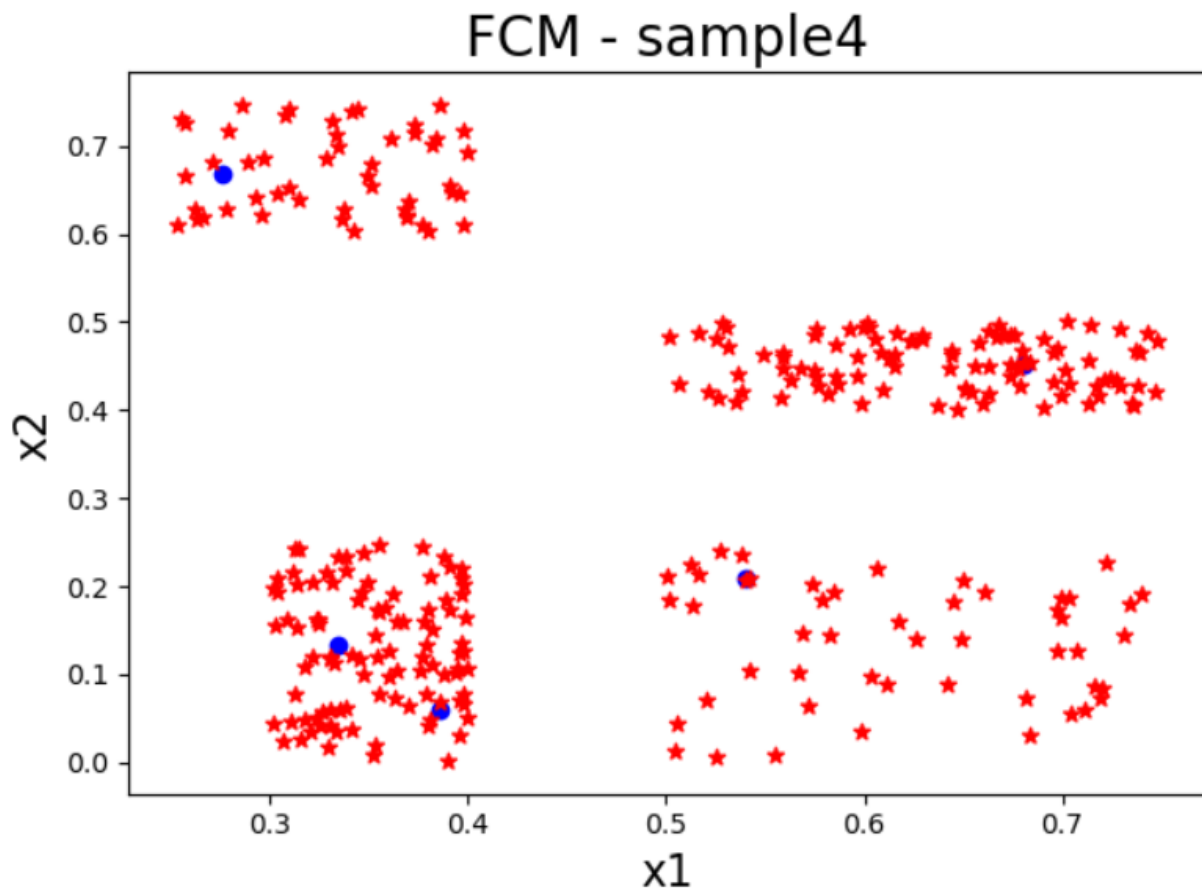
آنتروپی sample1 با تعداد خوشه ۴ مقدار 6.628 به دست آمد.



آنتروپی sample2 با تعداد خوشه 4 مقدار 7.209 به دست آمد.



آنتروپی sample3 با تعداد خوشه 3 مقدار 7.7219 به دست آمد.



آنتروپی sample4 با تعداد خوشه 5 مقدار 9.042 به دست آمد. یکی از این مراکز در گوشه سمت چپ بالا، دوتا گوشه سمت چپ پایین، یکی تقریباً در $x_1=0.55$, $x_2=0.2$ و دیگری در $x_1=0.7$, $x_2=0.45$ قرار دارد.