

برای فاز 2 این چند موضوع را یادم رفته بود اضافه کنم:

1. رویکرد اولیه سازی که در بالا ذکر شد از مراکز همزمان محافظت نمی کند (این امر به ویژه در هنگام برخورد با مجموعه داده هایی با نقاط تکراری مانند داده های پیکسل مشکل ساز است). اگر در پایان هر تکرار متوجه یک خوشه تک تنی (یعنی خوشه ای که تنها عضو آن مرکز آن است)، می توانید با پیدا کردن نقطه ای که بیشترین سهم را در خطای خوشه آن دارد، این مشکل را برطرف کنید. مرکز خوشه تک قلو باشد. به طور مشابه، اگر  $E > 1$  خوشه های تک تنی وجود دارد، نقاط E را که بیشترین سهم را در خطاهای خوشه های مربوطه دارند، بیابید و این نقاط را مراکز خوشه های تک تنی قرار دهید. برای آزمایش برنامه اصلاح شده خود، مجموعه داده های Iris Bezdek را با ایجاد مصنوعی 10 کپی از هر نقطه 5 داده تغییر دهید. مطمئن شوید که مجموعه داده های اصلاح شده ("iris\_bezdek\_mod.txt") دارای 420 نقطه داده (150 نقطه اصلی +  $150 * (1-10)$  نقاط تکراری) است. خود iris\_bezdek.txt را تغییر ندهید. یک برنامه کوچک جداگانه بنویسید که iris\_bezdek.txt را می خواند و iris\_bezdek\_mod.txt را خروجی می دهد. iris\_bezdek\_mod.txt را در آزمایشات و ارسال خود بگنجانید.

2. آیا دسته k-means که به طور یکنواخت به صورت تصادفی مقداردهی شده اند (مانند بالا) می توانند در هر تکرار با خوشه های خالی مواجه شوند؟ اگر بله، توضیح دهید که چه زمانی و چرا ممکن است این اتفاق بیفتد. در غیر این صورت توضیح دهید که چرا این امکان وجود ندارد. توضیحات خود را به عنوان نظر در بالای فایل منبع اصلی خود قرار دهید.

3. موردی را در نظر بگیرید که یک نقطه داده بیش از یک مرکز نزدیک داشته باشد. اکثر پیاده سازی ها چنین نقطه ای را به نزدیکترین مرکز با کوچکترین شاخص اختصاص می دهند. به نظر شما چرا چنین طرحی در ادبیات به یک قرارداد تبدیل شده است؟ اگر در عوض، چنین نقطه ای را به نزدیکترین مرکز به طور تصادفی انتخاب کنید، آیا الگوریتم همچنان همگرا است؟ اگر نه، چرا که نه؟ این استراتژی تخصیص تصادفی را به عنوان یک گزینه به برنامه خود اضافه کنید و رفتار همگرایی نظری آن را در یک نظر در بالای فایل منبع اصلی خود توضیح دهید.

4. بسیاری از راهنماهای عمومی یا خاص زبان در مورد شیوه های برنامه نویسی خوب در WWW وجود دارد. شما باید یک مورد مناسب را شناسایی کنید، URL آن را در بالای کد(های منبع) برنامه خود قرار دهید، و از آن در سراسر این پروژه استفاده کنید. به طور کلی، شما باید به مسائل ساختاری (به عنوان مثال، مدولار بودن) بیشتر از موارد زیبایی (به عنوان مثال، نامگذاری و قالب بندی) توجه کنید. به خاطر داشته باشید که اگر برنامه شما نادرست است، رعایت یا عدم رعایت شیوه های برنامه نویسی خوب اهمیتی ندارد. باید یک گزارش 1-2 صفحه ای بنویسید و راه حل خود را توضیح دهید (به عنوان مثال، ساختار داده ها و الگوریتم های مورد استفاده در هر مرحله).

جواب استاد برای فاز 1 که به نظرم باید متغیرها اصلاح شوند :

"K"، "R"، "N"، "D" نام متغیرهای بی معنی هستند. من از آنها در مشخصات فاز استفاده کردم زیرا این یک سند علمی است نه یک کد.