

روش حل مسئله

مساله مطرح شده، یک مساله خوشه‌بندی است و با توجه به نیازهای مساله و تعداد گروه‌های مشخص به یک الگوریتم Centroid Clustering نیازمندیم.

برای حل مسئله، از الگوریتم خوشه‌بندی k-means با چند مقدار اولیه استفاده کردیم. مراحل الگوریتم k-means به صورت زیر تعریف می‌شود:

۱. به تعداد خوشه‌ها، مراکز تصادفی انتخاب می‌شوند.

۲. خوشه‌بندی با مراکز انتخاب شده انجام می‌شود و نقاط به نزدیک‌ترین مراکز به عنوان خوشه اختصاص می‌یابند.

۳. مراکز جدید با محاسبه میانگین نقاط داخل هر خوشه محاسبه می‌شوند.

۴. تا زمانی که مراکز جدید با مراکز قبلی متفاوت باشد مراحل ۲ و ۳ تکرار می‌شود.

از آنجا که انتخاب تصادفی مراکز ممکن است به جواب‌های مختلفی منجر شود که الزاماً بهترین خوشه‌بندی را ارائه نمی‌کنند تغییرات زیر را در الگوریتم پایه اعمال کردیم:

✓ در انتخاب اولیه مراکز از الگوریتم k-mean++ استفاده کردیم. به این صورت که اولین مرکز به صورت تصادفی از میان داده‌ها انتخاب

می‌شود و سپس مراکز بعدی از میان داده‌های انتخاب نشده با یک احتمال انتخاب می‌شوند که این احتمال با فاصله نقاط با مرکز قبلی

نسبت مستقیم دارد. در نتیجه نقاط با فاصله بیشتر احتمال بیشتری برای انتخاب مرکز بعدی دارند.

✓ در کد موجود الگوریتم k-mean به دفعات با مقادیر اولیه متفاوت اجرا می‌شود و خوشه‌بندی با کمترین امتیاز به عنوان پاسخ نهایی

خروجی داده می‌شود. که امتیاز هر خوشه‌بندی با مجموع فواصل نقاط هر خوشه از مرکز خوشه محاسبه می‌شود.

خروجی کد

در نهایت خروجی در فایل اکسل ذخیره شده و همچنین نقشه خوشه‌ها به صورت زیر نمایش داده می‌شوند:

