توضیح سوال 1:

در این سوال باید الگوریتم kmeans را اجرا کنیم. برای این کار ابتدا نقاط را که در یک فایل متنی هب صورت مختصاتی داده شند را میخوانیم و رسم میکنیم. . برای آزمایش نتایج را اول با استفاده ازتوابع آماده پایتون بدست آوردم و با نام تست ذخیره کردم.

برای پیاده کردن این الگوریتم، ابتدا یک سری نقطه رندوم به عنوان مرکز در نظر میگیریم. سپس برای هر نقطه دیگر فاصله را تا این نقاط میسنجیم و هر نقطه را به نزدیکترین هسته تخصیص میدهیم. سپس برای همه نقاط یک کلاستر، باید با میانگین گیری بین مختصات همگی آنها، یک مرکز جدید انتخاب کنیم. در واقع در کل این الگوریتم همواره 2 فاز دارد:

1) یافتن مراکز
2) تخصیص دهی مجدد به نزدیکترین کلاستر

چون مراکز خوشه ها در ابتدا به طور رندوم تعیین میشوند، ما با هر بار اجرای این الگوریتم نتیج متفاوتی میگیریم اما در کل به طور خوبی نقاط تقسیم میشوند و هر یک در خوشه درستی قرار میگیرند.

لازم به ذکر است چون اینجا تعداد نقاط کم بود و نیز الگوریتم پیچیدگی زمان بالایی ندارد، من یک معیار نسبتا دقیقی روی کلاستر کردن نقاط گذاشتم. اگر کلاسترها تغییر نکردند، یا در واقع به جایی رسیدیم که فاصله مرکز جدید یک کلاستر با مرکز در مرحله قبلی برای آن کلاستر از 0.0000001 کمتر بود، و این برای همه نقاط برقرار بود، آنگاه الگوریتم پایان میابد.

لازم به ذکر است که بار اول کلاسترینگ را در فضای با بردار های ویژگی x وyانجام دادیم و لذا همواره برحسب این 2 معیار کلاسترینگ انجام میشود ولی هرگز نمیتوان با استفاده از معادله خط 2 بعدی در این فصا، به خوشه بندی ای رسید که وقتی یک خوشه درون دیگری است ، آنها را از هم جدا کند. لذا برای اینکه اینگونه نیز کلاستربندی کنیم، باید به فضای قطبی برویم. در این فصا بر حسب زاویه و شعاع برای نقاط بردار ویژگی تشکیل میدهیم و سپس سایر مراحل برای کلاستر کردن مانند قبل است. تنها ورودی kmeans به جای x وy نقاط، theta و r است. این تبدیل مختصات نیز در تابع cast_to_polar_coordinaton انجام شده است.
نتایج نیز در تابع show_results با دو رنگ متفاوت برای کلاسترها نمایش داده شدند. همچنین مرکز هر کلاستر نیز در شکلها نمایش داده شده است.