



بسمه تعالی
 دانشگاه صنعتی شریف
 دانشکده علوم کامپیوتر
 Image processing – دکتر مصطفی کمالی
 HW4 q1 report

سوال 1 – Kmeans :

برای این سوال یک تابع به نام `kmean` داریم که نقاط و تعداد `cluster`ها را میگیرد و خوشه بندی میکند و مرکز خوشه ها و همینطور اعضای خوشه ها را خروجی میدهد ، به اینصورت که ابتدا به صورت رندم به تعداد `k` مرکز خوشه یکتا از ارایه ورودی برمیدار ، سپس برای تمام نقاط موجود در `arr` ورودی ، بررسی میکند که هر کدام از این نقاط به کدام یک از مراکز خوشه ها نزدیک تر است ، سپس `label` آن نقطه را برابر مرکز خوشه نزدیکتر قرار میدهد و به `cluster` آن مرکز میپیوندد . پس از اینکه برای تمام نقاط پیمایش کردیم ، مرکز جدید هر خوشه برابر میانگین نقاط داخل آن است .

حال اگر تفاضل نسبی مرکز کلاستر جدید و قبلی از مقدار `epsilon` برابر `0.0001` بیشتر بود ، مرکز خوشه بروز شده و برابر مقدار جدید قرار میگیرد و الگوریتم آنقدر تکرار میشود که هیچ کدام از مراکز خوشه ها از جای خود کمتر از `epsilon` تکان بخورند ، در اینصورت الگوریتم همگرا شده و مراکز خوشه ها و اعضای هر خوشه را خروجی میدهد.

در ابتدا برای نقاط ورودی مسیله با استفاده از `plt.scatter` رسم میکنیم که با میتوان دید که در دو خوشه نقاط قرار میگیرند ، ولی با اجرای الگوریتم `kmeans` چون مراکز اولیه خوشه ها رندم تعیین میشود و همینطور داده ها در `feature space` فعلی جدایی پذیری خوبی ندارند لذا هر بار نتیجه نسبتا متفاوتی بدست میاید.

حال اگر نقاط را از فضای فعلی به مختصات قطبی ببریم و بر اساس مختصات (ρ, θ) رسم کنیم خواهیم دید که دو گروه کاملا از هم جدا شده اند که $\rho = \sqrt{x^2 + y^2}$ و $\theta = \arctan(y/x)$. سپس با استفاده از این `feature` های جدید که `map` شده `feature` های قبلی اند عمل `clustering` را `k=2` و روش `kmeans` هر چندبار که انجام دهیم خواهیم دید که نتیجه یکسان خواهد بود ، فقط ممکن است که `label` هر خوشه بسته به انتخاب نقاط اولیه عوض شود که ایرادی ندارد زیرا موقعیت مراکز خوشه و نقاطی که در یک خوشه قرار میگیرند ثابت میماند.