

بنام خدا

تمرین چهارم

درس داده کاوی

- به زبان برنامه نویسی پایتون برنامه ای بنویسید که موارد خواسته شده زیر را انجام دهد.
- موارد خواسته شده را به دقت انجام دهید.
- از خروجی های بدست آمده در هر مرحله عکس بگیرید و در فایل پاسخ قرار دهید. پاسخ های خود را به طور کامل به صورت صوتی شرح دهید.
- فایل نهایی توضیحات به همراه کدها و خروجی های بدست آمده را به صورت یک فایل فشرده (rar یا zip) در لینک زیر آپلود کنید. اگر ویدیو قرار داده می شود بایستی حجم آن تا حد امکان کاهش داده شود.
- به کسی که بهترین و کاملترین محتوا برای هر تمرین را تولید کند (هم از نظر صحت پاسخ و هم از نظر گرافیکی و هم ارزش آموزشی برای دیگران) نمره اضافه نیز در نظر گرفته خواهد شد.

<https://cloud.vru.ac.ir/s/wnNGrNfFeo6SH3H>

خوشه بندی داده ها

چهار مجموعه داده دو بعدی در فایل های "spiral.txt"، "flame.txt"، "jain.txt" و "Aggregation.txt" قرار دارند که به همراه تمرین برای شما ارسال شده اند.

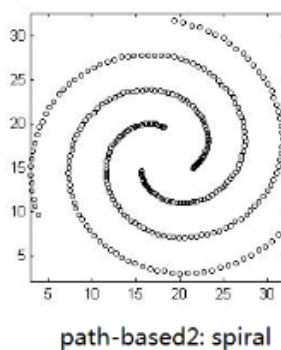
در هر فایل در هر سطر برای هر نمونه داده سه مقدار به صورت زیر قرار دارد.

مقدار بعد اول (X)	مقدار بعد دوم (Y)	شماره خوشه های مشخص شده
31.95	7.95	3
31.15	7.3	3
30.45	6.65	3
29.7	6	3
28.9	5.55	3
28.05	5	3

برای هر کدام از این مجموعه داده ها موارد زیر را انجام دهید.

۱. بدون در نظر گرفتن شماره خوشه های داده شده در فایل، پراکندگی داده ها را در یک نمودار نمایش دهید.

مانند نمودار زیر:



۲. بدون در نظر گرفتن شماره خوشه های داده شده در فایل، داده های دو بعدی را با روش خوشه بندی k-means خوشه بندی کنید و مجدد داده ها را در یک نمودار نمایش دهید به طوری که داده های هر خوشه به رنگ متفاوتی نمایش داده شوند.

K یعنی تعداد خوشه ها را برای هر مجموعه داده از ستون سوم فایل داده بدست آورید.

راهنمایی: `from sklearn.cluster import KMeans`

۳. بدون در نظر گرفتن شماره خوشه های داده شده در فایل، داده های دو بعدی را با روش خوشه بندی سلسله مراتبی خوشه بندی کنید و مجدد داده ها را در یک نمودار نمایش دهید به طوری که داده های هر خوشه به رنگ متفاوتی نمایش داده شوند.

راهنمایی: `from sklearn.cluster import AgglomerativeClustering`

۴. بدون در نظر گرفتن شماره خوشه های داده شده در فایل، داده های دو بعدی را با روش خوشه بندی dbscan خوشه بندی کنید و مجدد داده ها را در یک نمودار نمایش دهید به طوری که داده های هر خوشه به رنگ متفاوتی نمایش داده شوند.

راهنمایی: `from sklearn.cluster import DBSCAN`

راهنمای بیشتر: از لینک زیر کمک بگیرید

<https://nzlul.medium.com/clustering-method-using-k-means-hierarchical-and-dbscan-using-python-5ca5721bbfc3#:~:text=DBSCAN%20can%20identify%20clusters%20in,points%20are%20noise%20or%20outliers.>

موفق باشید