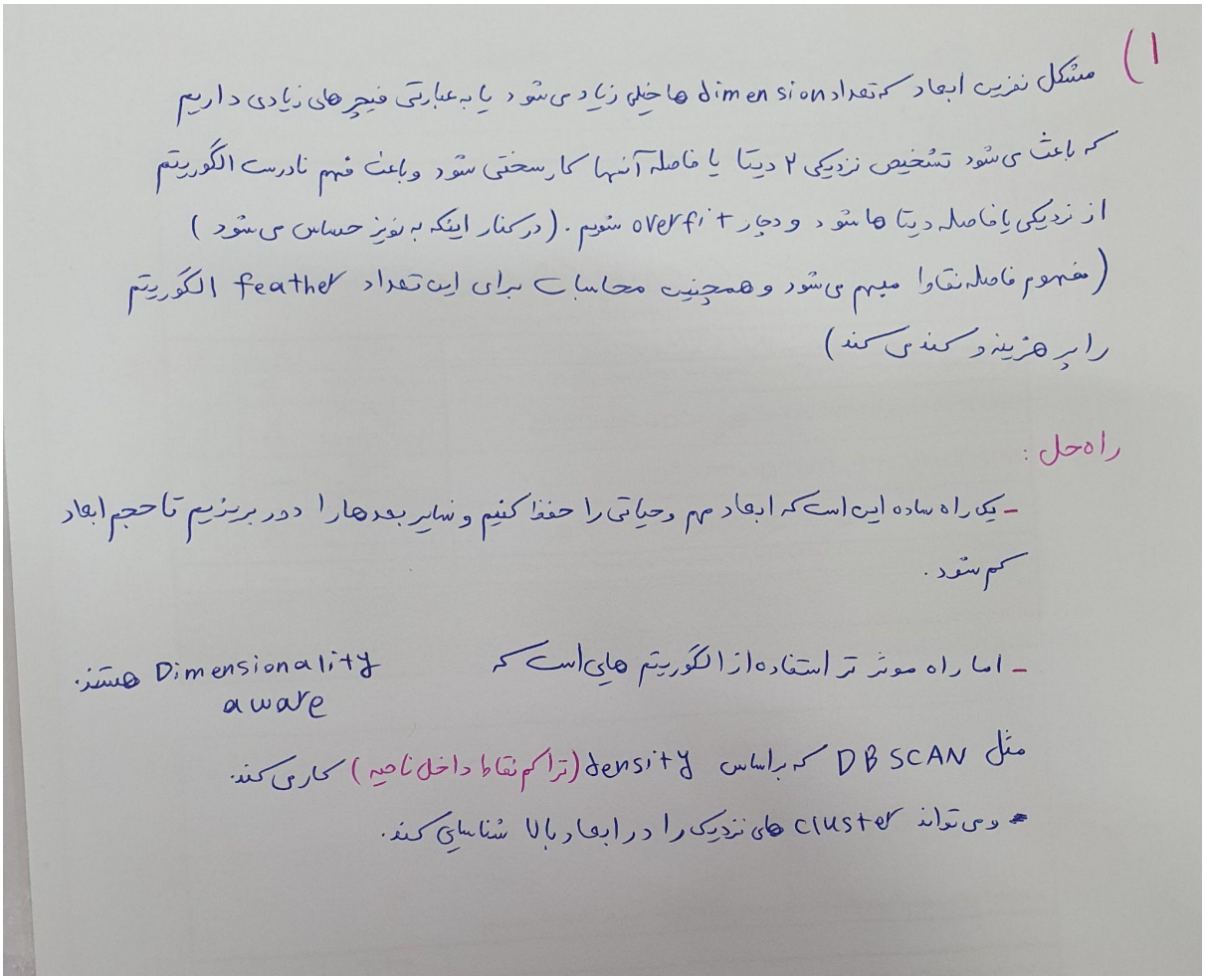
تمرین چهارم

آرمان غفارنیا 99243056

سوالات تحلیلی :

1)



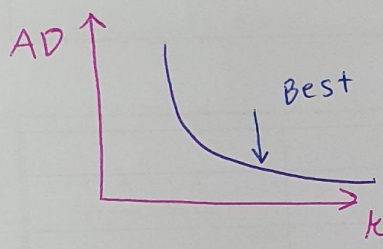
یك راه حل موثر دیگر استفاده از تجزیه ماتریس SVD است
راستاهایی که بیشترین شدت تغییرات داده در آنها اتفاق افتاده است را حفظ میکنیم و سایرین را حذف میکنیم که باعث کاهش ابعاد میشود و تقریب خوبی با خطای کمینه به ما میدهد.

(2و3)

۲) یک چالش مهم انتخاب k یا تعداد خوشه ها است .

یک راه که در اسلاید مطرح شده این است که یک رنج از k ها را تست کنیم و میانگین فاصله نقاط با centroid را برای هر k بدست بیاوریم.
نقطه میانگین

طبیعتاً با افزایش تعداد خوشه ها این عدد کم می‌شود اما از یکجایی به بعد این کاهش چشمگیر نیست .
این نقطه ، نقطه مناسب k است .



شکل اسلاید .

۳) برای انجام فشرده سازی به کمک k-means باید یک فرآیند clustering انجام دهیم

به این صورت که رنگ هایی که در یک رنج هستند را در یک cluster قرار دهیم و مرکز خوشه را به عنوان نماینده آن رنگ فرض کنیم .

$k \rightarrow$ تعداد cluster ها

تعداد رنگ هایی که در حالت فشرده ذخیره می کنیم

برای انجام clustering ، با دریافت تصویر ورودی که شامل پیکسل هاست هر پیکسل را یک نقطه فرض می کنیم که شامل ۳ تا فیچر R و G و B است . که به کمک آنها الگوریتم را اجرا می کنیم .

(4

۴) DBSCAN ، زیرا داده‌های موجود در شکل حالت تودرتو دارند (nesting)

و احتمالاً سایر الگوریتم های clustering دچار مشکل می‌شوند. و دسته بندی نادرست می کنند،

مثلاً k-means که براساس فاصله اقلیدسی نقاط کار می کند طبیعتا نمی‌تواند نقاط آبی پررنگ

را به درستی دسته بندی کند. (چون در حداکثر فاصله هستند یکسری از نقاط آبی)

اما الگوریتم DBSCAN که براساس Density کار می کند، می‌تواند داده‌های تودرتو را

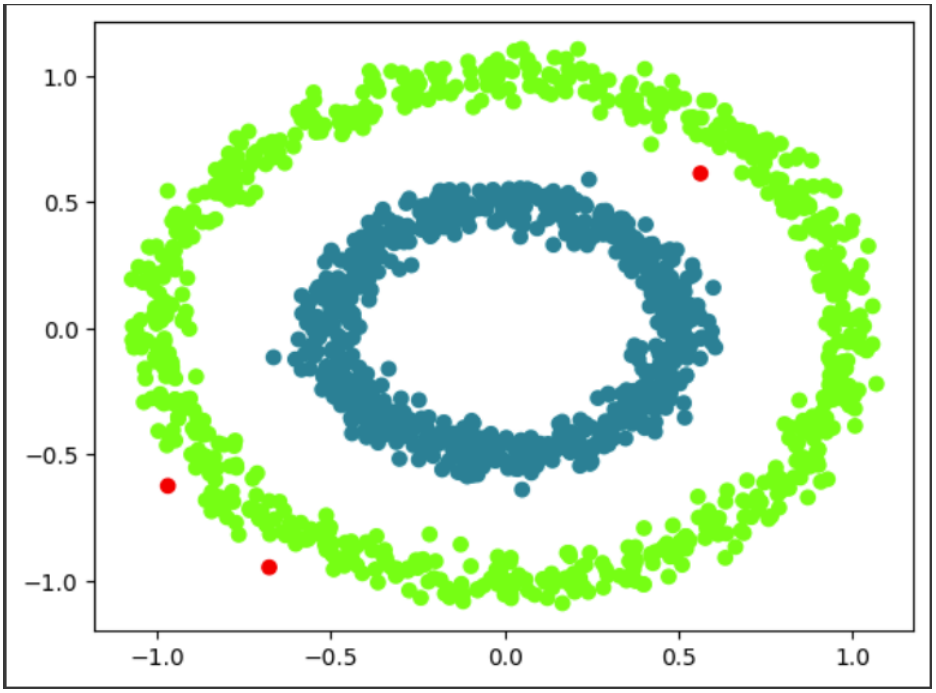
تراکم نقاط
در ناحیه

دسته بندی مناسب کند.

سوال کدی :

1)

خروجی دیتاست اول :



نقاط قرمز noise هستند.

خروجی دیتاست دوم :