



دانشگاه یزد

گروه مهندسی کامپیوتر

رشته تحصیلی: هوش مصنوعی و رباتیکز

نام درس: یادگیری ماشین

تکلیف شماره ۴

استاد مربوطه: دکتر مهدی یزدیان

تهیه کننده: حمیدرضا نادمی

بهار ۱۳۹۸

A. Distance-Based (k-means)

۱. مدل k-mean را پیاده سازی کنید.

فایل main.py ، خط ۳۹ الی ۶۹

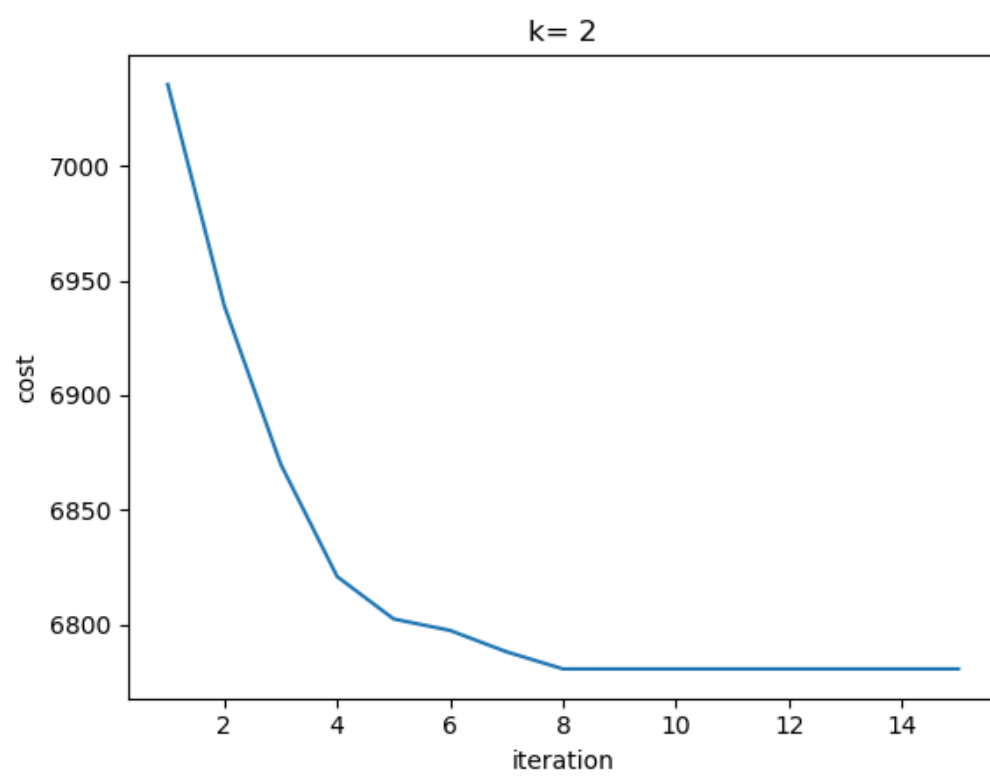
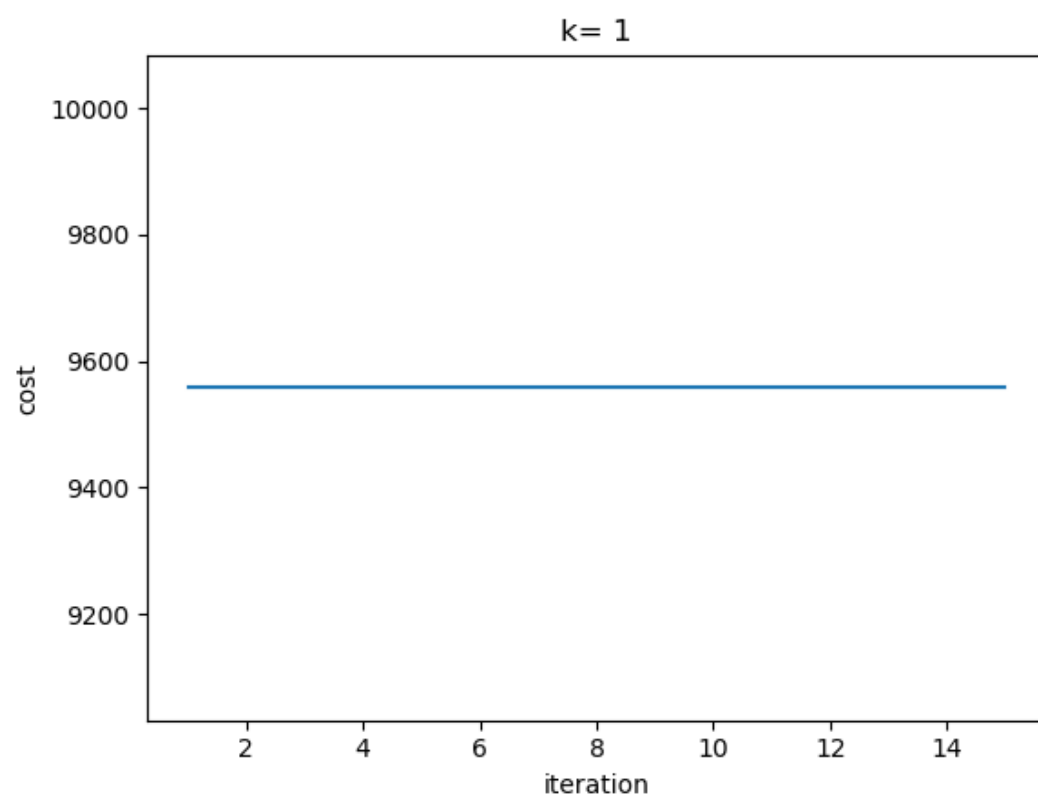
۲. به ازای $k = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ الگوریتم را بر روی داده ها اجرا کنید.

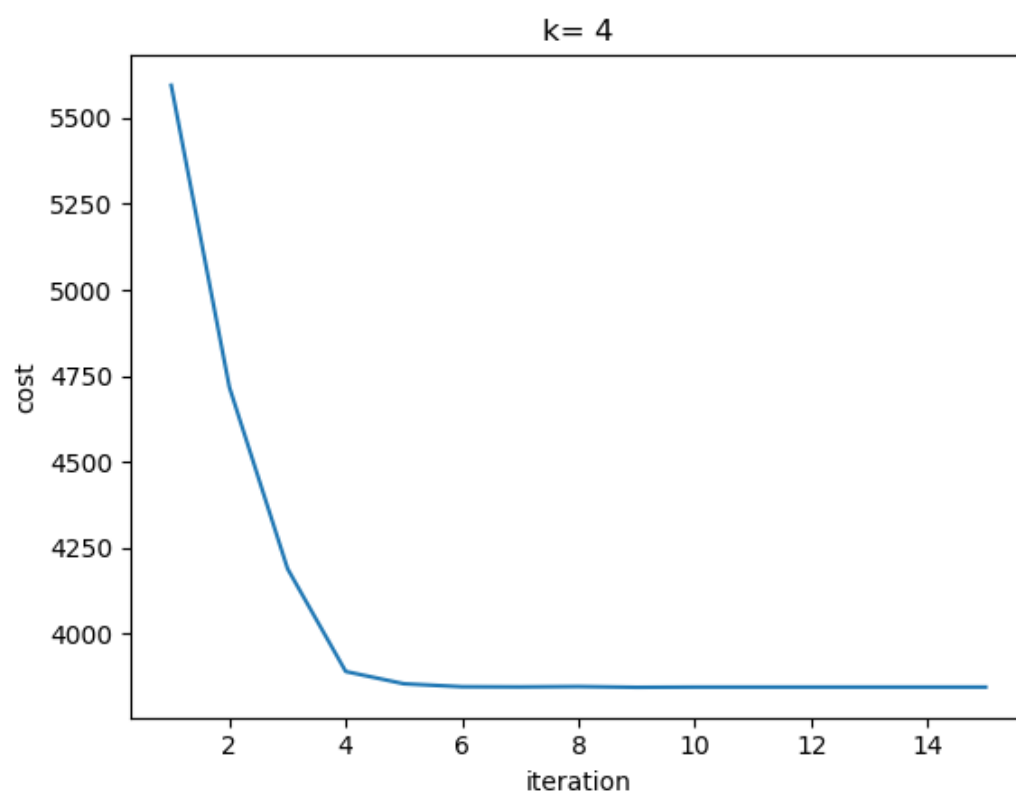
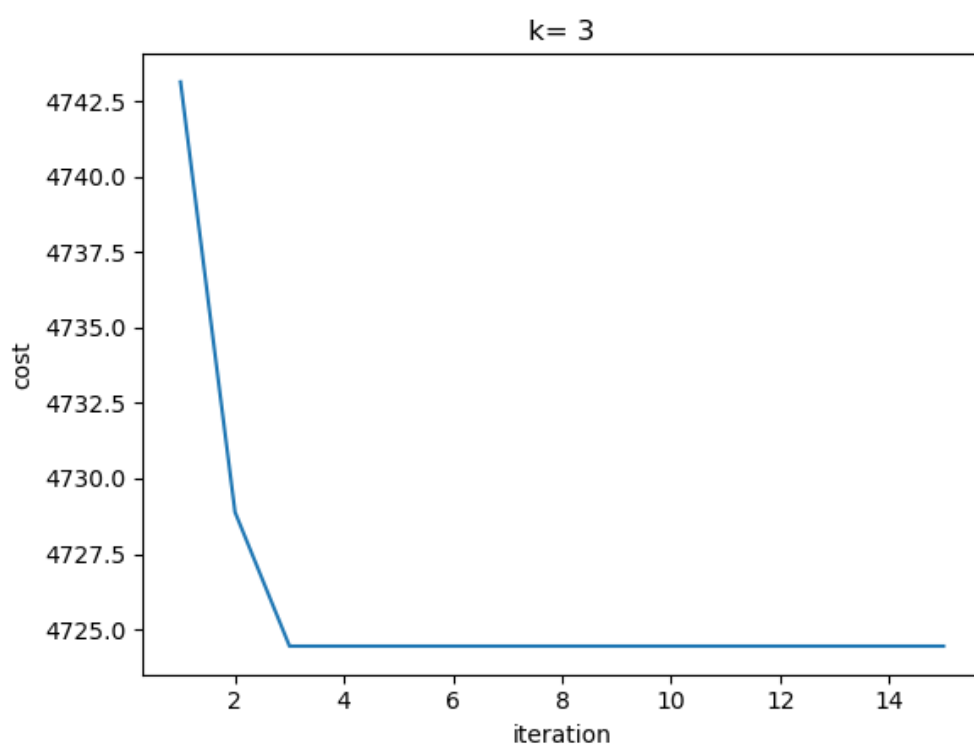
فایل main.py ، خط ۱۱۴ الی ۱۲۹

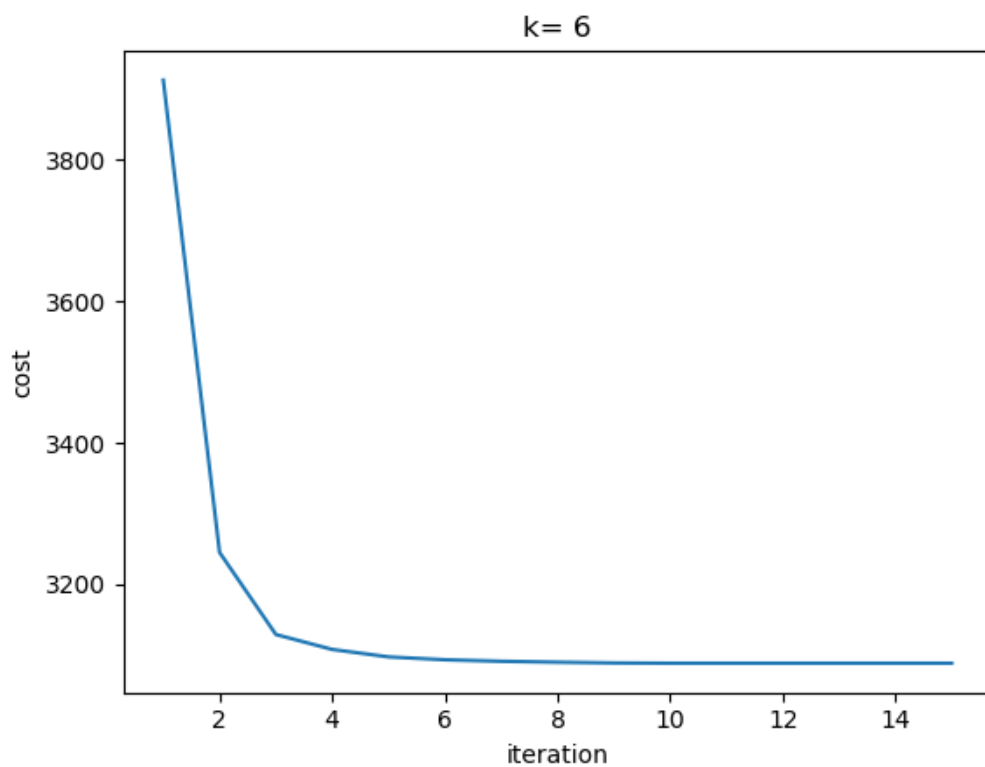
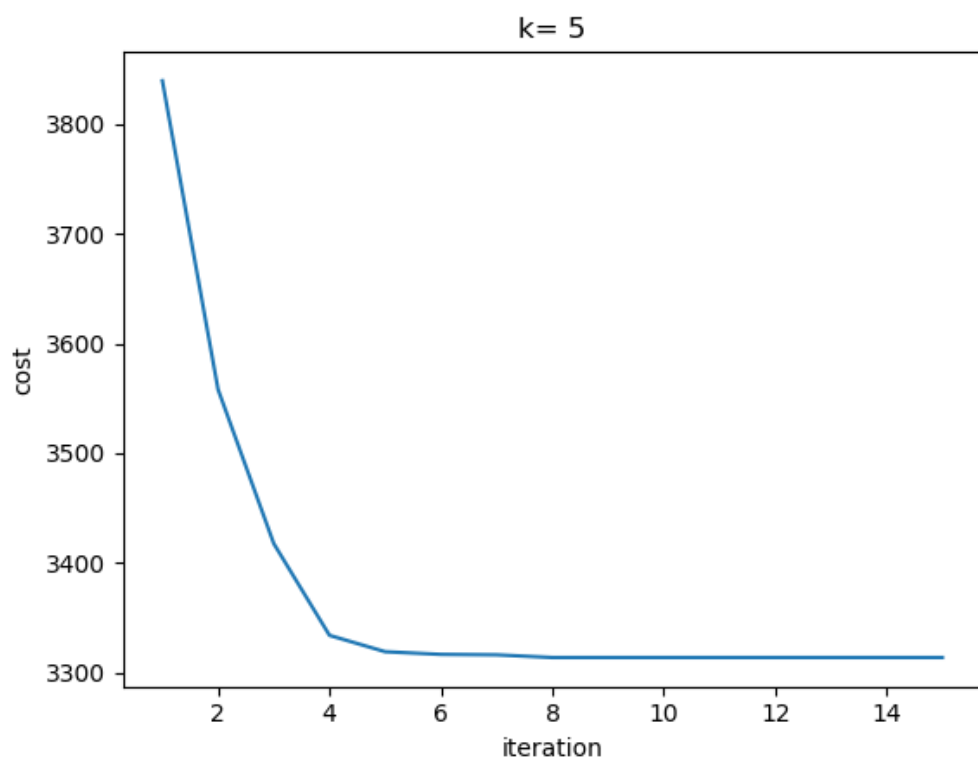
۳. تابع هزینه به ازای هر ۱۰ حالت قسمت قبل را بدست بیاورید.

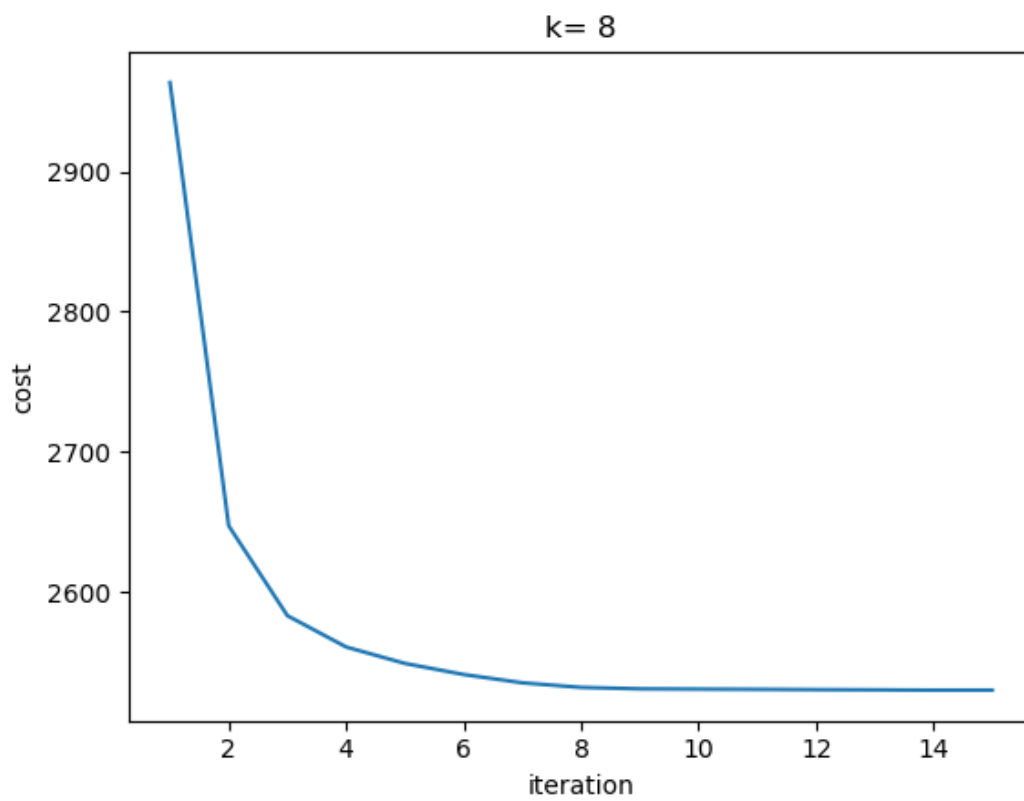
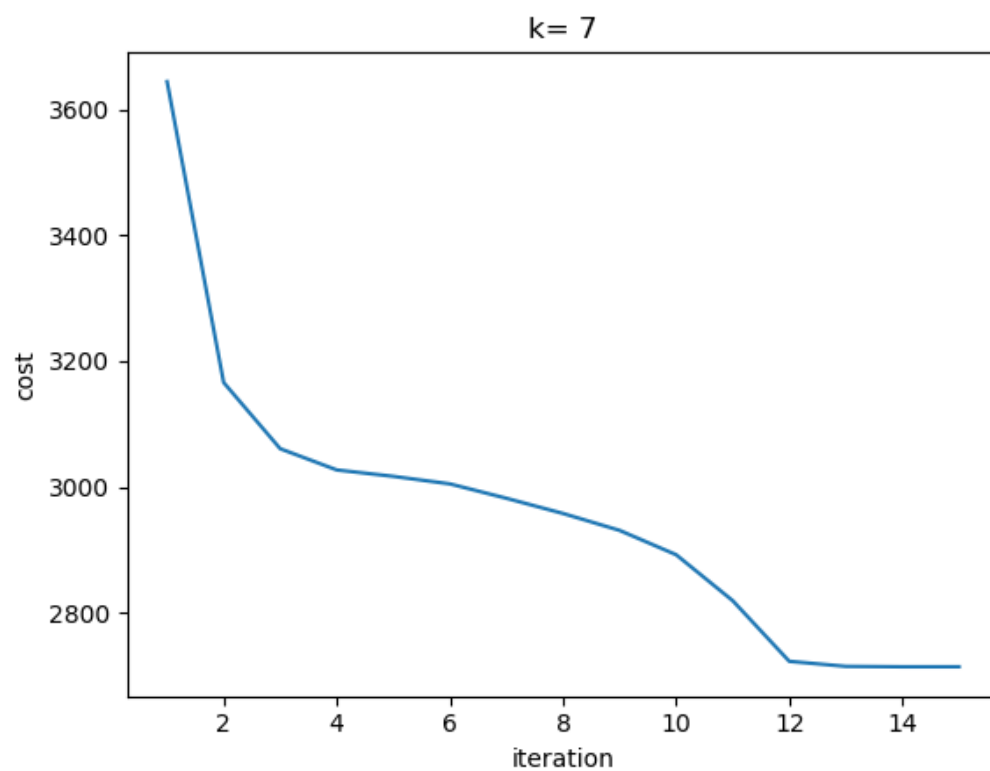
به ازای هر گام مقدار تابع هزینه محاسبه شده است، حداکثر تعداد گام های رسیدن به همگرایی ۱۵ در نظر گرفته شده است، با افزایش گام ها مقدار هزینه k-means کمتر می شود.

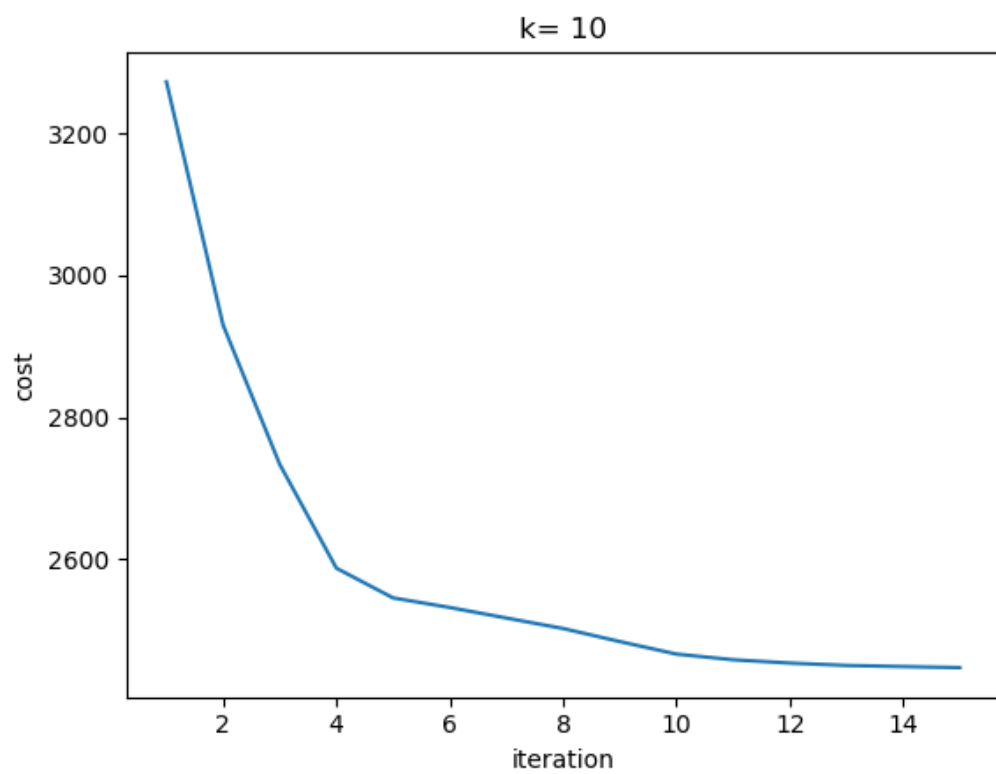
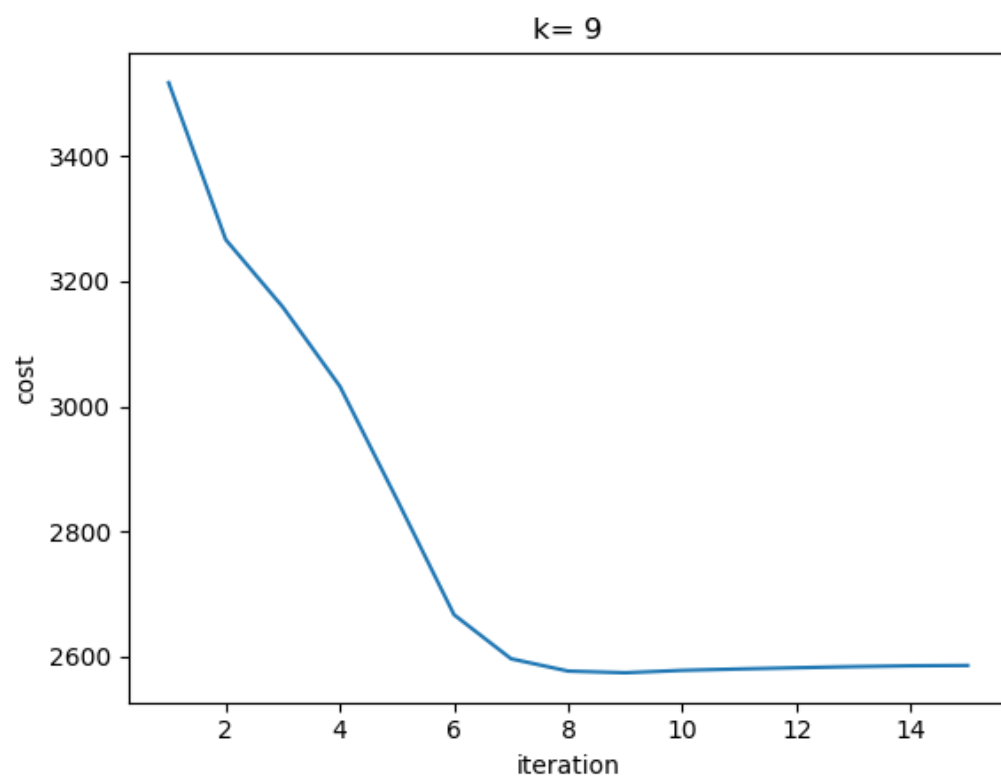
۴. نمودار تابع هزینه متناظر با قسمت قبل را رسم کنید.









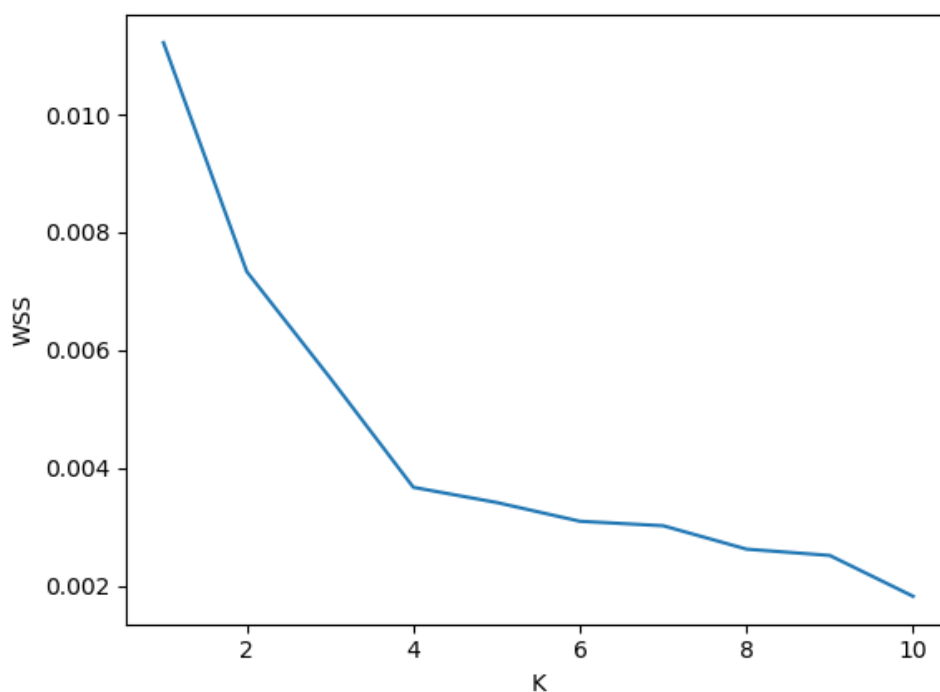


۵. روشی به منظور ارائه k بهینه پیشنهاد داده، آن را پیاده سازی و k بهینه در این مسئله را بدست آورید.

روش Elbow method پیشنهاد می‌شود که مراحل پیاده سازی آن به شرح زیر است:

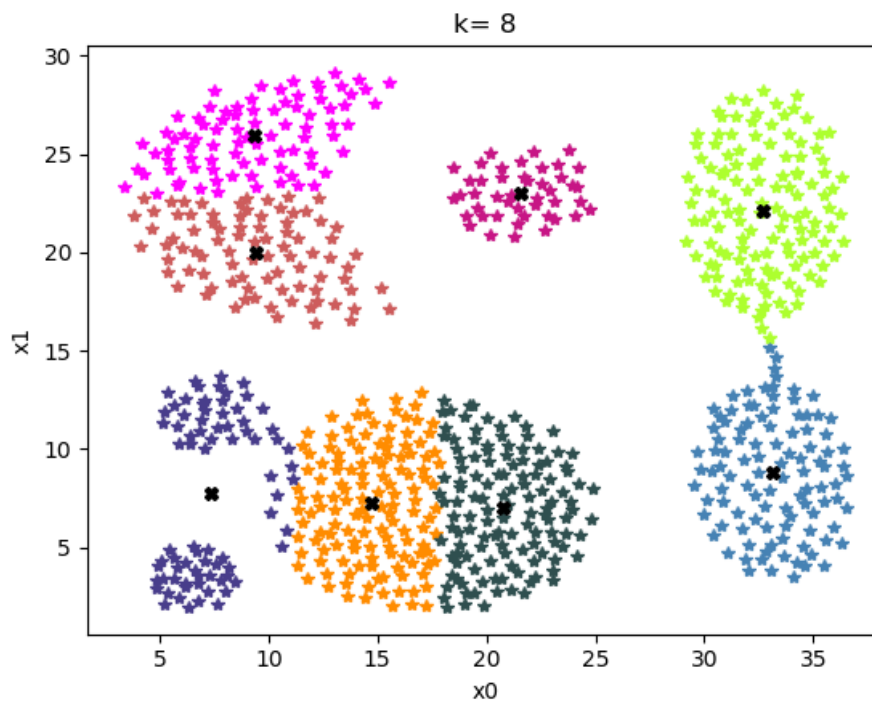
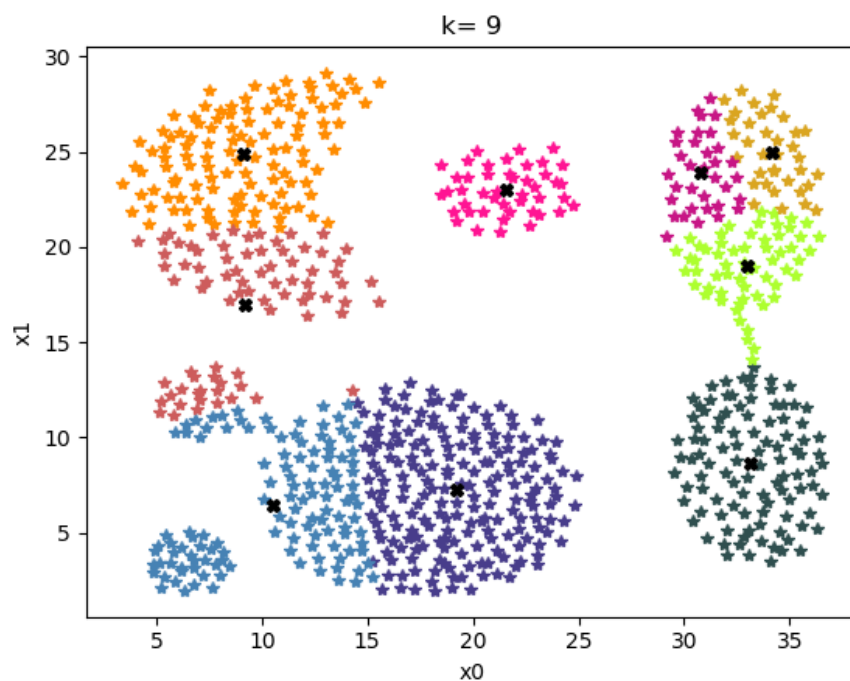
1. Compute clustering algorithm (e.g., k-means clustering) for different values of k . For instance, by varying k from 1 to 10 clusters.
2. For each k , calculate the total within-cluster sum of square (wss).
3. Plot the curve of wss according to the number of clusters k .
4. The location of a bend (knee) in the plot is generally considered as an indicator of the appropriate number of clusters.

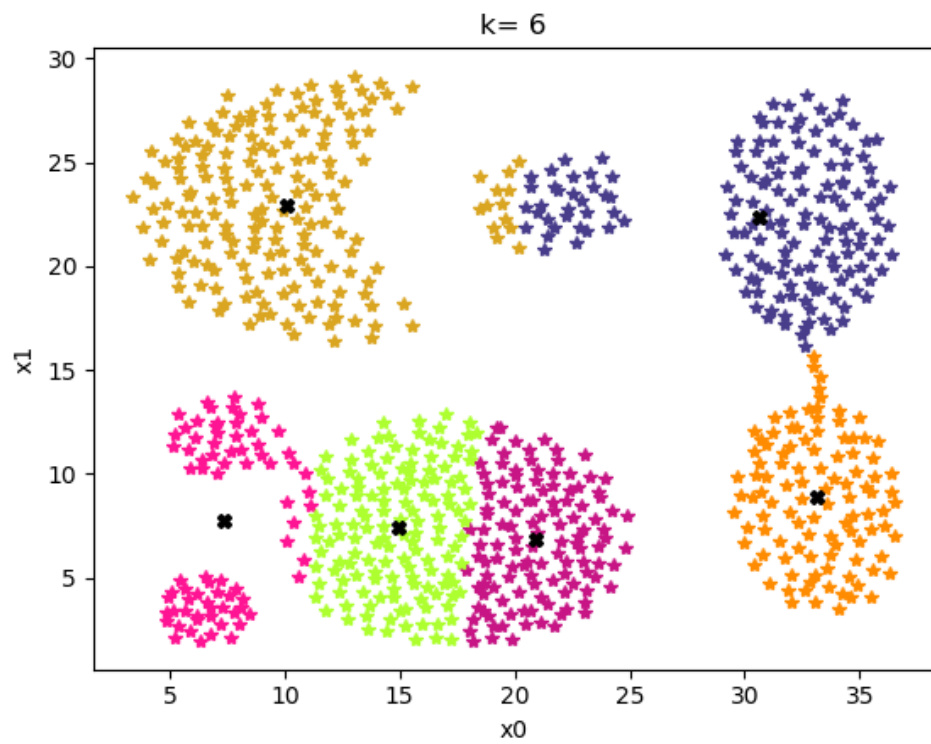
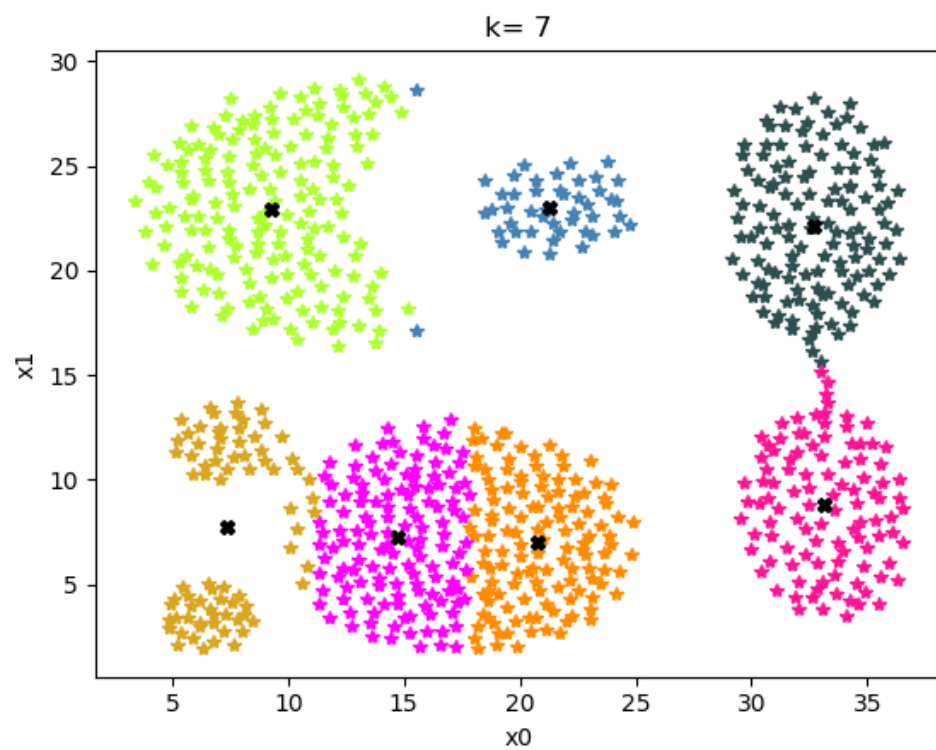
تابع `compute_wss` خط ۸۱ الی ۸۹ بدین منظور پیاده سازی شده است، عکس واریانس هر خوشه محاسبه شده و مقدار بدست آمده برای k خوشه باهم جمع می‌شود و امتیاز برای k -means در نظر گرفته می‌شود، برای ۱۰ حالت ممکن k این مقدار محاسبه می‌شود و با رسم نمودار k و امتیاز متناظر آن، نقطه آرنج در نمودار به عنوان k بهینه در نظر گرفته می‌شود.

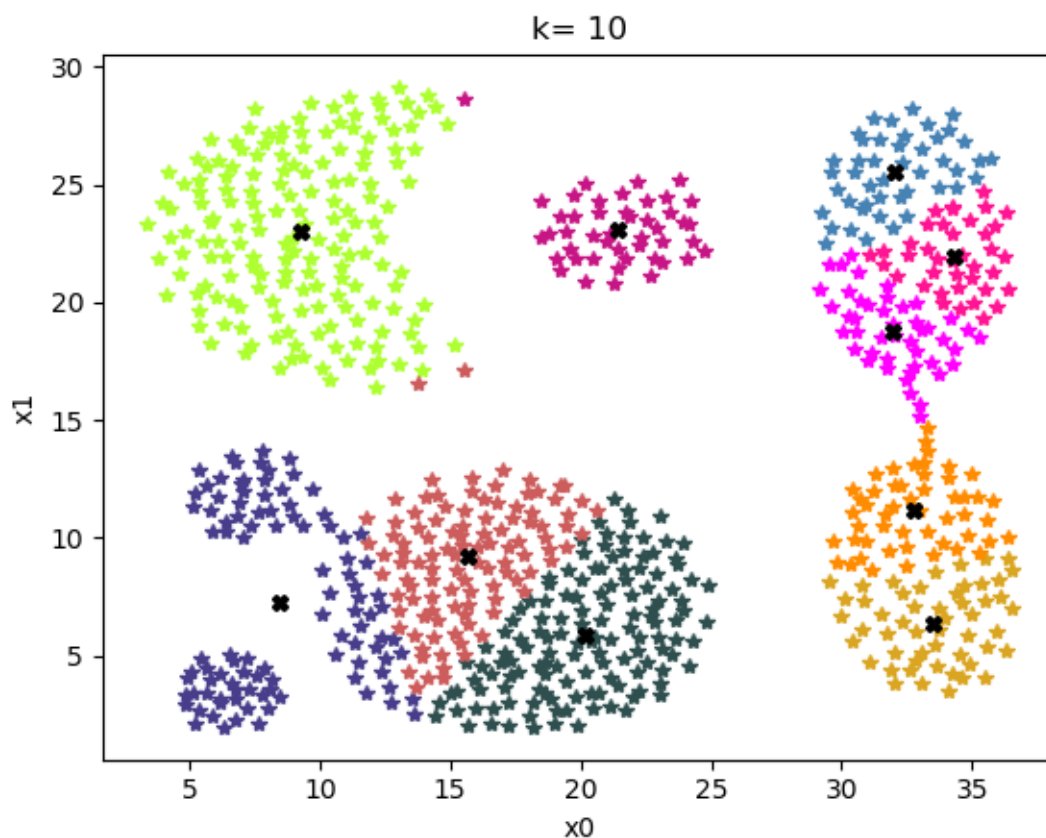


مقدار k برابر ۶، بهینه ترین مقدار برای خوشه بندی به روش K-means است.

۶. از ۱۰ مدل بدست آمده در قسمت ۲، شکل حاصل از نتیجه کلاسنبدی توسط ۵ مدل با دقت بهتر را رسم کنید.







B. Density-Based (DBSCAN)

۱. به انتخاب خودتان، حداقل ۱۰ مقدار متفاوت برای MinPts و ۱۰ مقدار متفاوت برای ϵ در نظر بگیرید.

فایل part_B.py، خط ۱۸ و ۱۹

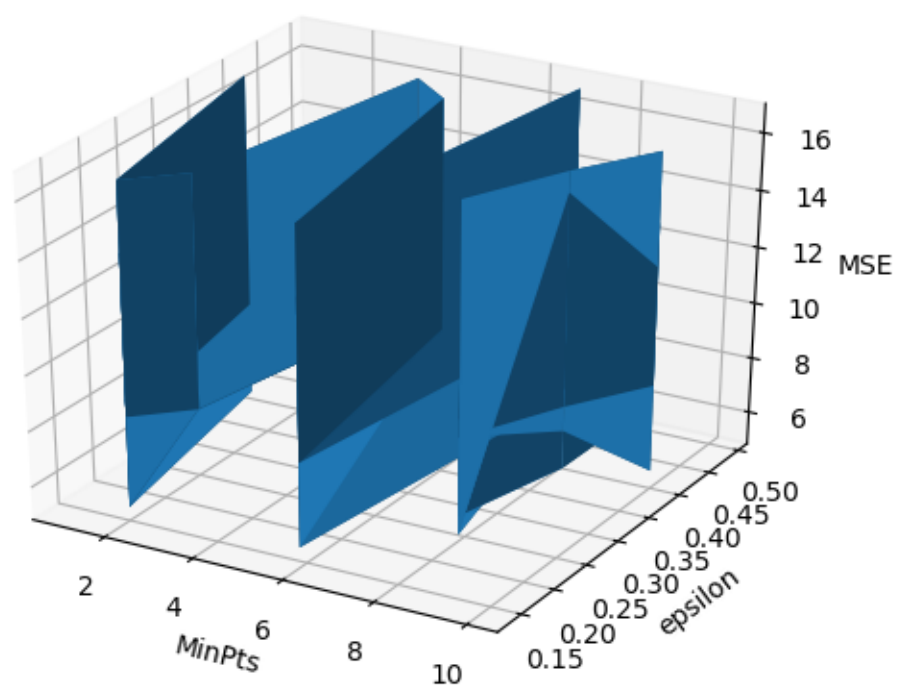
۲. مدل DBSCAN، متناظر با مقادیر دو به دو پارامترها در قسمت قبل را، بر روی داده ها اجرا کنید.

(توجه کنید حداقل $10 \times 10 = 100$ مدل ایجاد می شود)

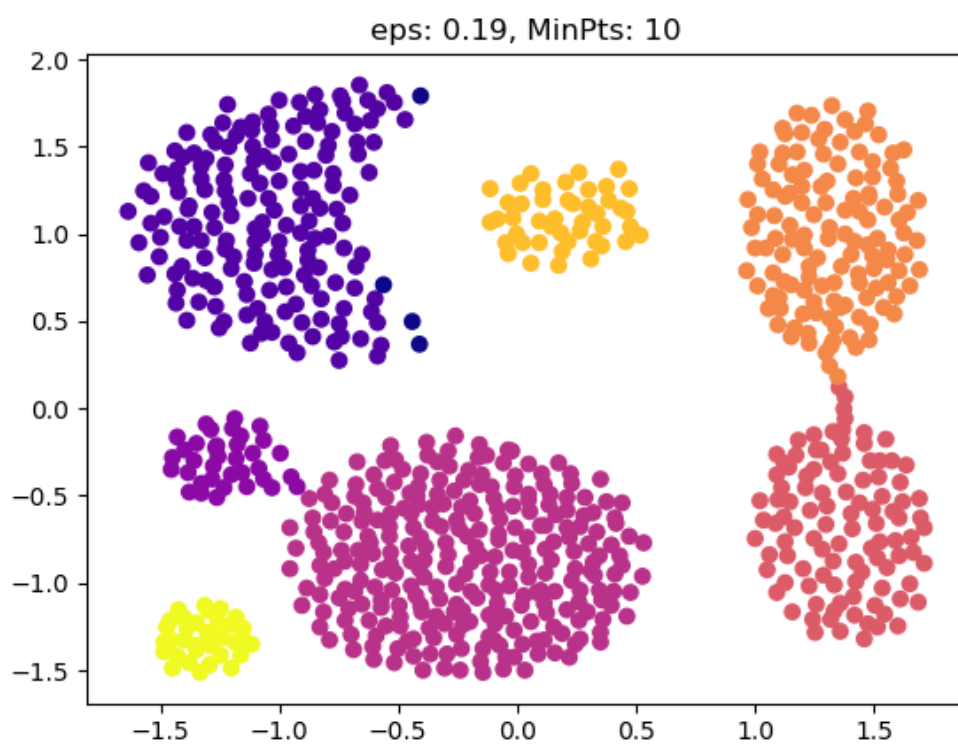
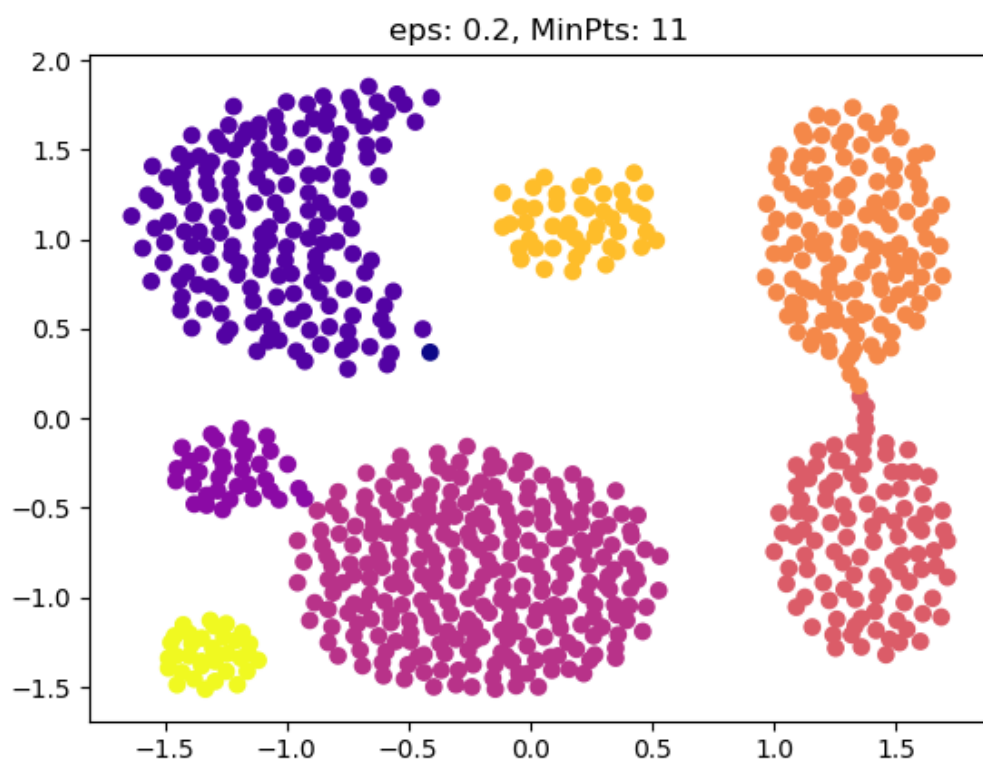
فایل part_B.py، خط ۲۶ الی ۳۴

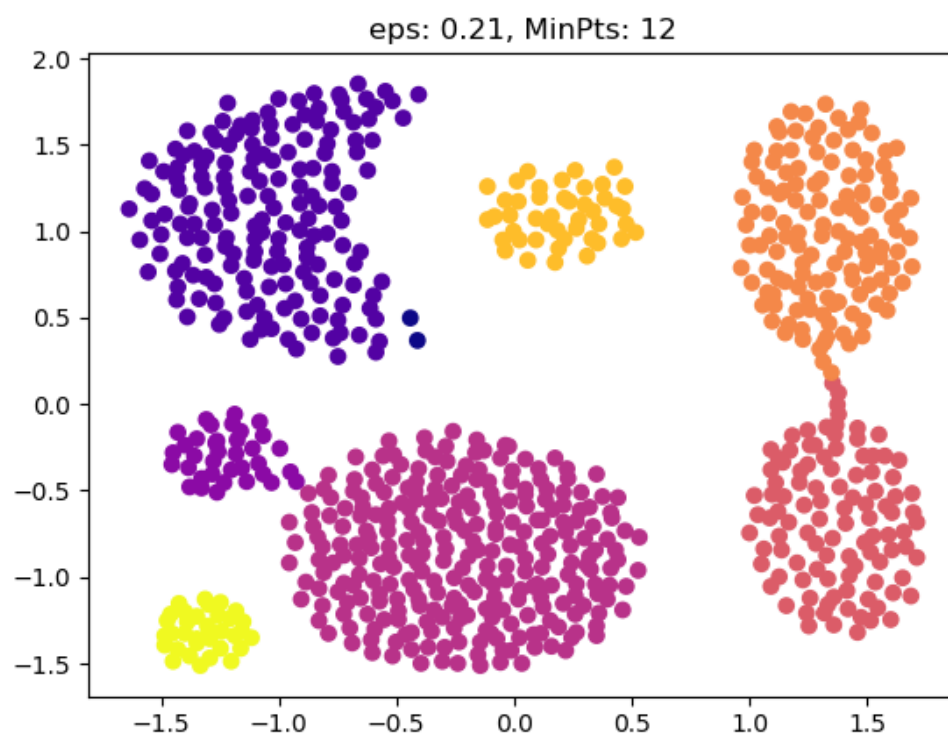
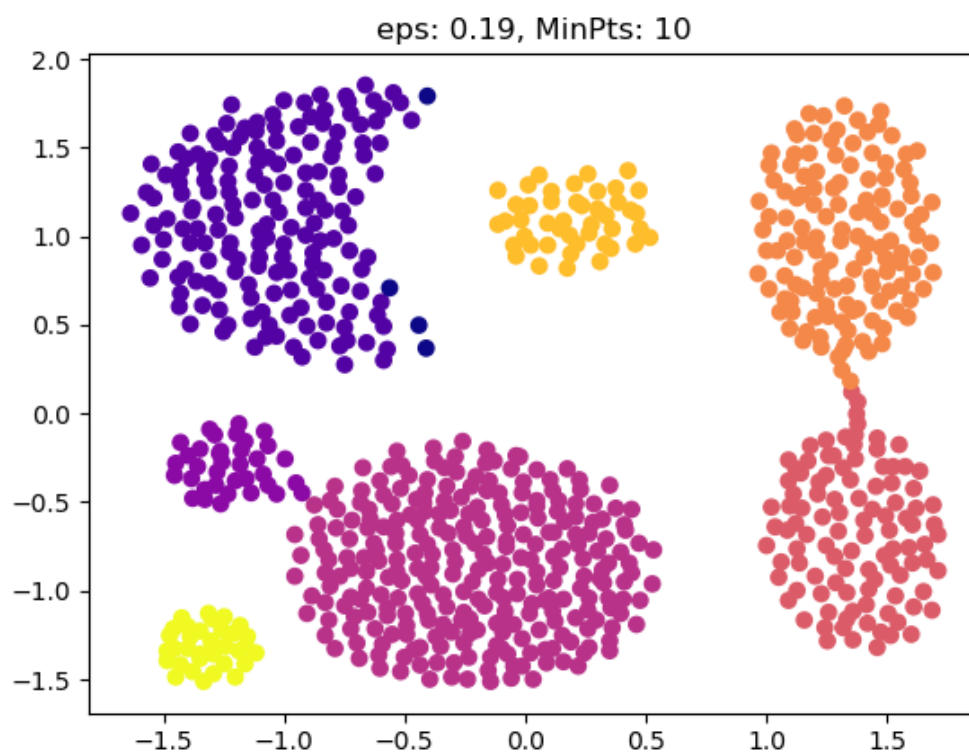
۳. با استفاده از روش mean squared error (MSE)، نمودار MSE متناظر با مدل های بدست

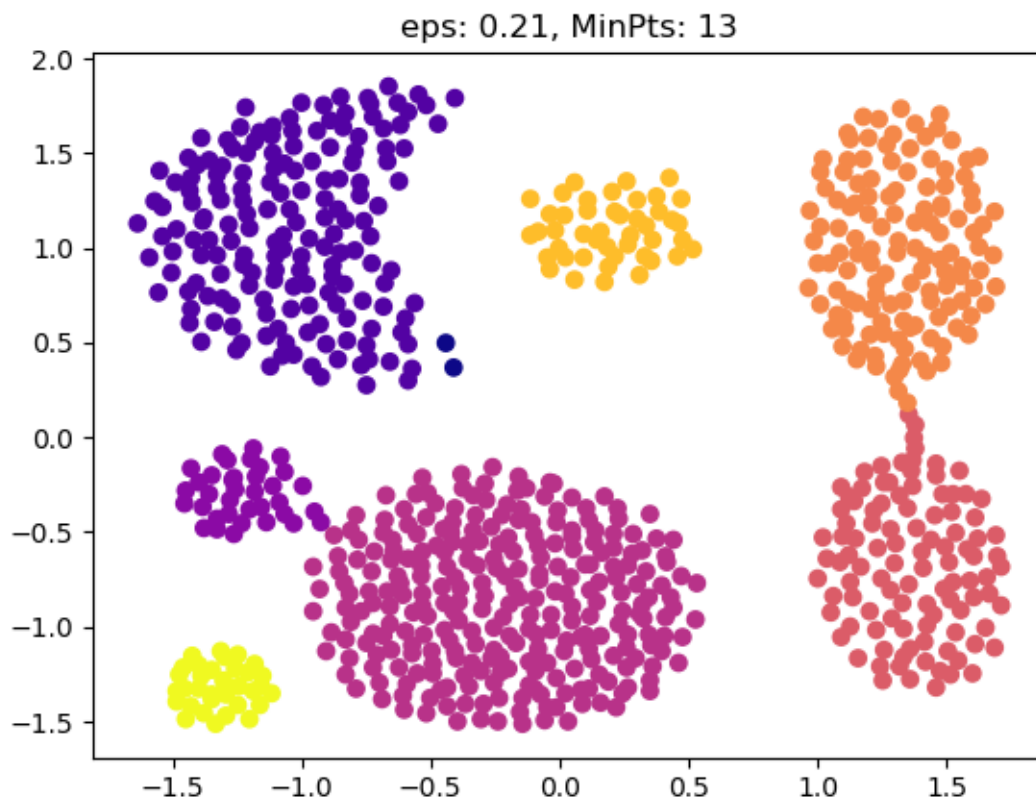
آمده در بخش قبل را به ازای پارامترهای مختلف رسم کنید. (نمودار سه بعدی: MinPts، ϵ ، MSE)



۴. از مدل های بدست آمده در بخش ۲، شکل حاصل از نتیجه کلاسیکندی توسط ۵ مدل با دقت بهتر را رسم کنید.







۵. نتیجه بدست آمده در بخش قبل را با آنچه در بخش ۶ از تمرین A بدست آمد مقایسه کنید و نتایج حاصل را تحلیل کنید.

- با انتخاب min point و epsilon مناسب میتوانیم با دقتی خیلی خوب نسبت به k-means داده‌ها را خوشه بندی کنیم
- K-means نسبت به DBSCAN سریعتر است.
- در DBSCAN برای خوشه بندی نیازی به مشخص کردن تعداد خوشه ها نداریم