

به نام خدا
یادگیری ماشین (بهار ۱۳۹۸)
استاد: آقای دکتر مهدی یزدیان
کمک استاد: خانم فرزانه نیک زاد
پروژه چهارم: *Unsupervised Learning*

در این تمرین از دیتاست Aggregation که پیوست تمرین شده است، استفاده کنید.

A. Distance-Based (k-means)

۱. مدل k-mean را پیاده سازی کنید.
۲. به ازای $k = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ الگوریتم را بر روی داده ها اجرا کنید.
۳. تابع هزینه به ازای هر ۱۰ حالت قسمت قبل را بدست بیاورید.
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.
۶. از ۱۰ مدل بدست آمده در قسمت ۲، شکل حاصل از نتیجه کلاسدی توسط ۵ مدل با دقت بهتر را رسم کنید.

توجه: لازم است مدل k-means و تابع هزینه آن را خودتان پیاده سازی کنید.

B. Density-Based (DBSCAN)

۱. به انتخاب خودتان، حداقل ۱۰ مقدار متفاوت برای MinPts و ۱۰ مقدار متفاوت برای ϵ در نظر بگیرید.
۲. مدل DBSCAN، متناظر با مقادیر دو به دوی پارامترها در قسمت قبل را، بر روی داده ها اجرا کنید.
(توجه کنید حداقل $10 * 10 = 100$ مدل ایجاد می شود)
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 کلاسدی 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.

توجه: به منظور استفاده از مدل DBSCAN میتوانید از توابع آماده پایتون استفاده کنید.

موفق باشید