

دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

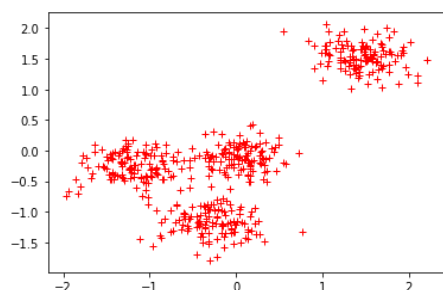
گزارش تمرین سری چهارم داده کاوی

استاد دکتر مزلقانی

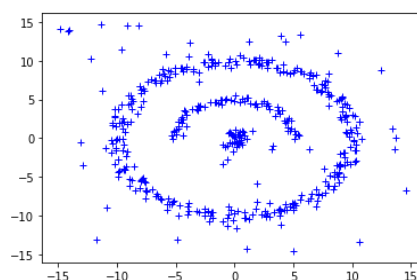
محبوبه شاکری ۹۵۳۱۳۰۱

تمرین اول Kmeans :

الف) داده ها رو لود میکنیم و به شکل زیر نمایش میدهم



dataset1



dataset2 ۱ Figure

۲ Figure

الف)

خوشه بندی برای $K=2,3,4$ به این صورت خواهد

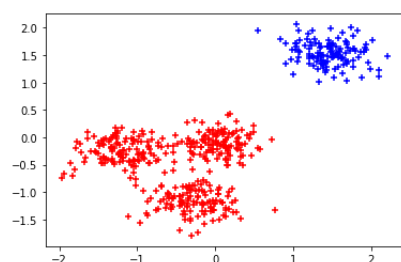


Figure 3 $k=3$

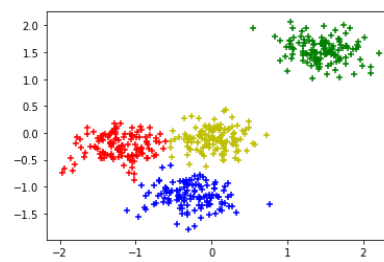


Figure 4 $k=2$

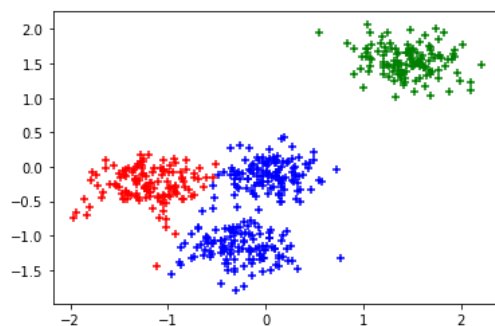


Figure ۵ $k=3$

ب) میزان خطای هر خوشه به شرح زیر می‌باشد.

$K=2$ [0.7491160326963786, 0.6408392294166209]

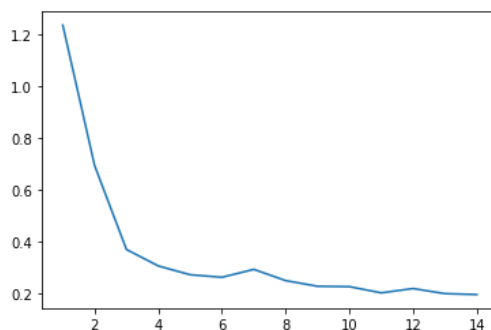
$K=3$ [0.3160088195773474, 0.3346341775026478, 0.4913832583180853]

$K=4$ [0.2886597397595568, 0.30638902552400693, 0.31630045732935297, 0.3162275478913516]

ج) به ترتیب برای K از صفر تا ۱۵ مقدار خطای خوشه بندی به شکل زیر است

[1.236496530958413, 0.6949776310564997, 0.3806754184660268, 0.30689419262606704,
0.2738629082548584, 0.2678108009593281, 0.2574670598472235, 0.24535751011357043,
0.24901387380917342, 0.22859412649297317, 0.2290870388049254, 0.21500354781073372,
0.19714192647996365, 0.2069551244236427]

د)



شکل ۱ خطای خوشه بندی

ه) با توجه به نقطه ی زانویی نمودار بهترین K برابر با ۴ است.

ی) به دلیل اینکه این داده های این دیتاست دارای واریانس و میانگین تقریباً برابری هستند این الگوریتم نمیتواند مرکز مناسبی پیدا کند و همگرا نمیشود

و همینطور که مشخص است خوب دسته بندی نمیشوند و دارای خطای خوشه بندی بالایی هستند

[8.658345154205232, 6.960987232813734, 5.373815377701159, 4.997880671835755,
3.9714149142051967, 3.505004270619317, 3.138702433982107, 2.837217207521079,
2.588096795273229, 2.4776300527042965, 2.2323244593448175, 1.9039898533228292,
2.0774559984930616, 1.7551662928641332]

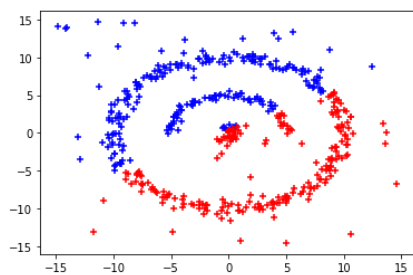


Figure 6 $K=4$

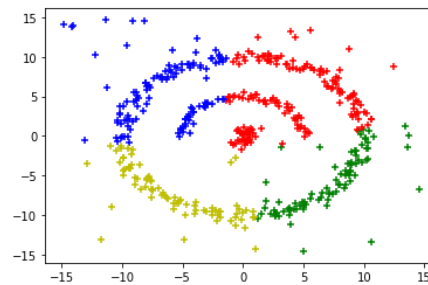


Figure 7 $K=2$

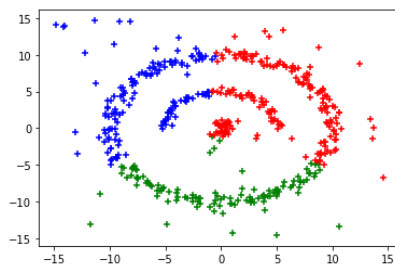
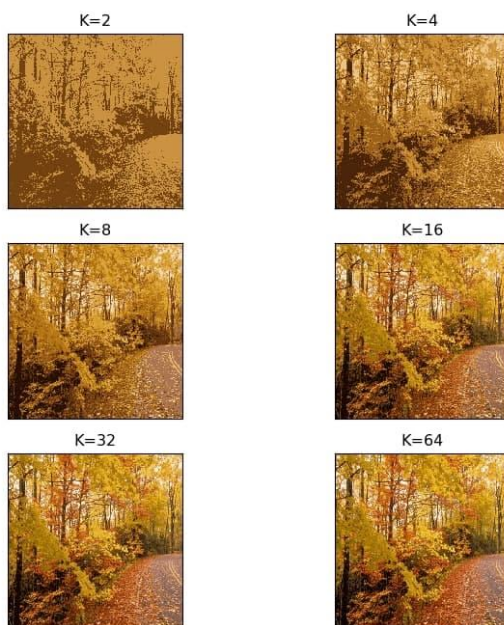


Figure8 $k=3$

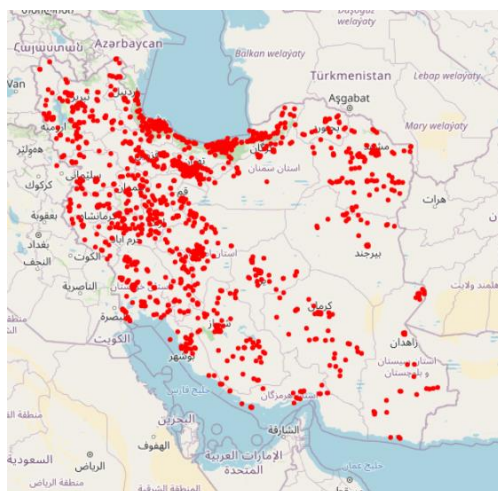
(۲)



Compressed Images^۷ Figure

(۳)

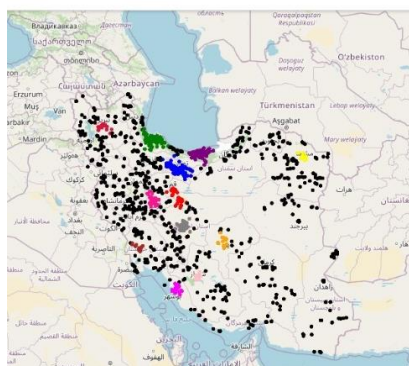
ابتدا تمام نقاطی که کرونا در آنها رویت شده را نشان میدهیم.



ب) الگوریتم DBSCAN را با مقادیر دیفالت خودش $\text{eps}=0.5$ و $\text{minPoints}=5$ اجرا میکنیم.

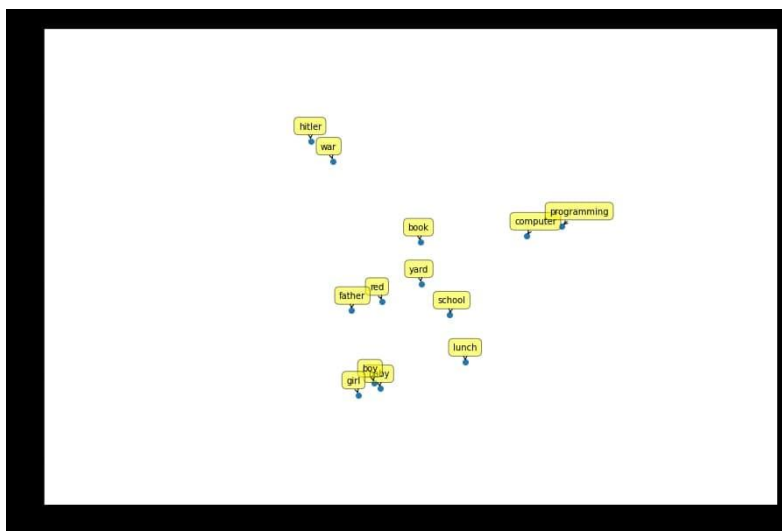


ج و د)

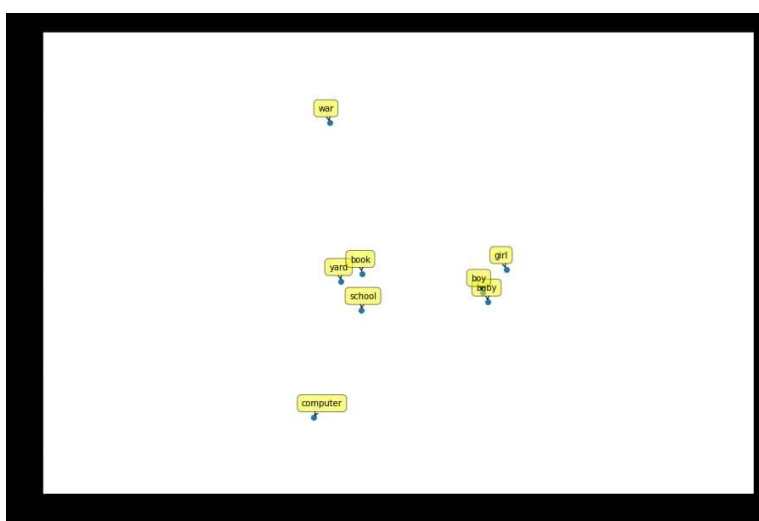


(۴)

کلمات نزدیک بهم شباهت بیشتری دارند



شکل ۲ نمونه کلمه



شکل ۳ نمونه کلمه