



دانشگاه صنعتی امیر کبیر
(پلی تکنیک تهران)

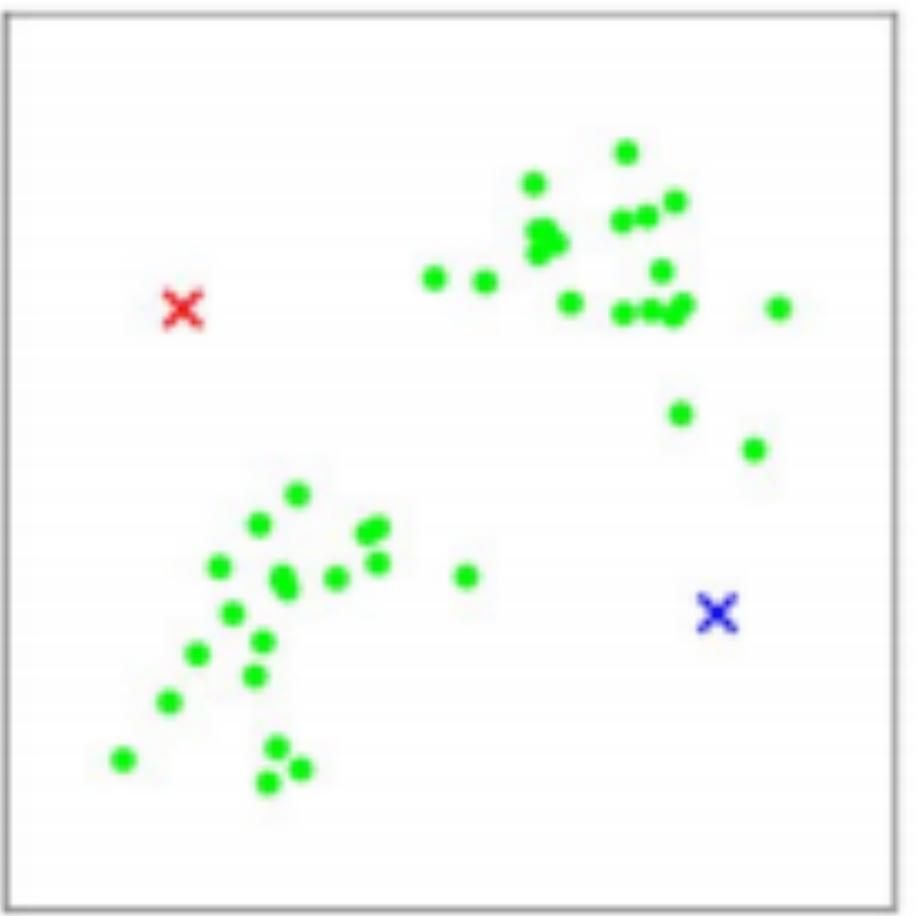
آزمایشگاه هوش محاسباتی آزمایش چهارم (پیاده‌سازی الگوریتم K-means)

پاییز ۱۴۰۱

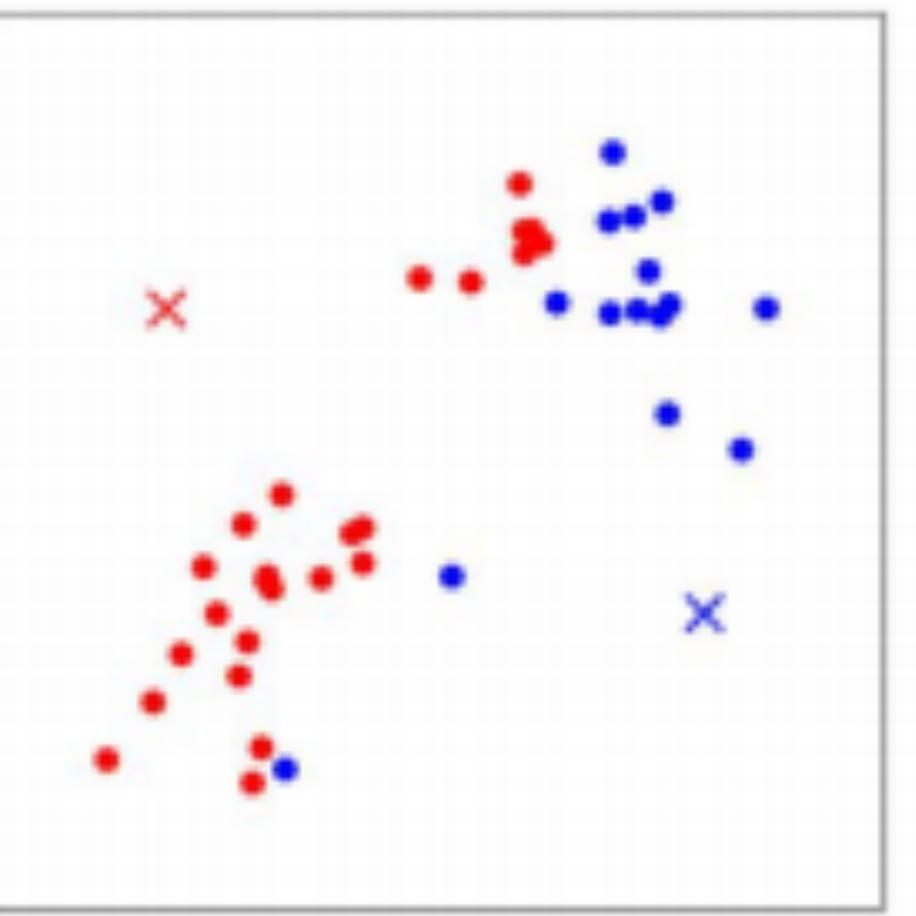
K-means الگوريتم



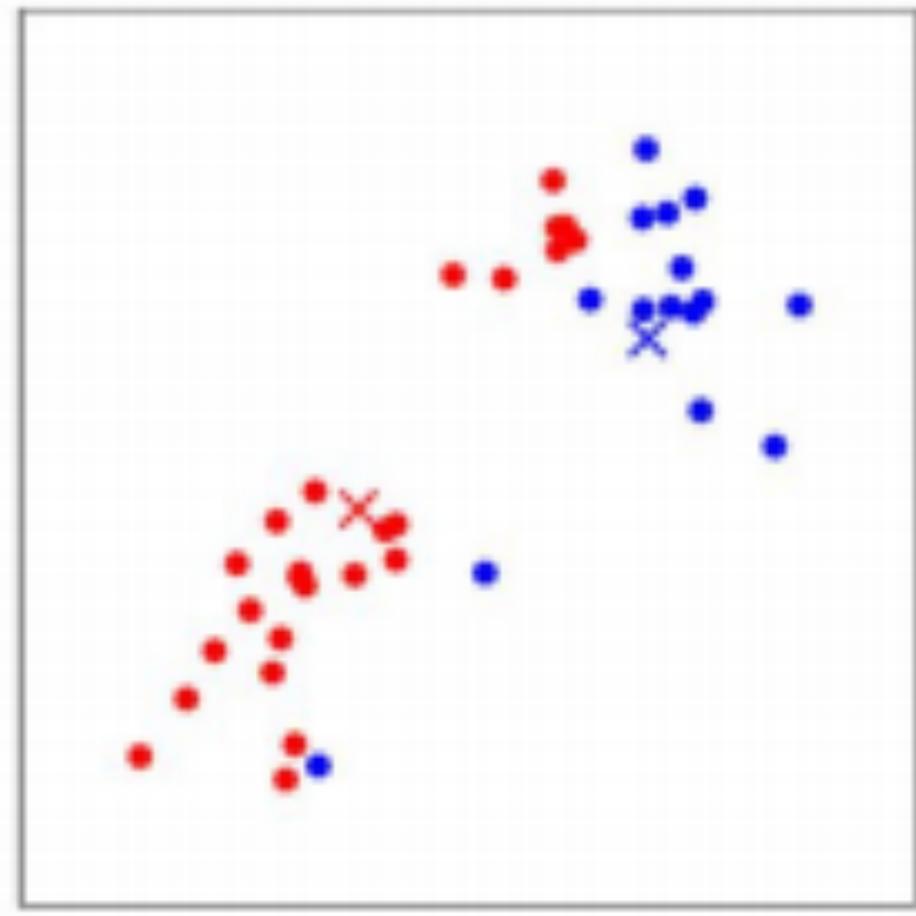
(a)



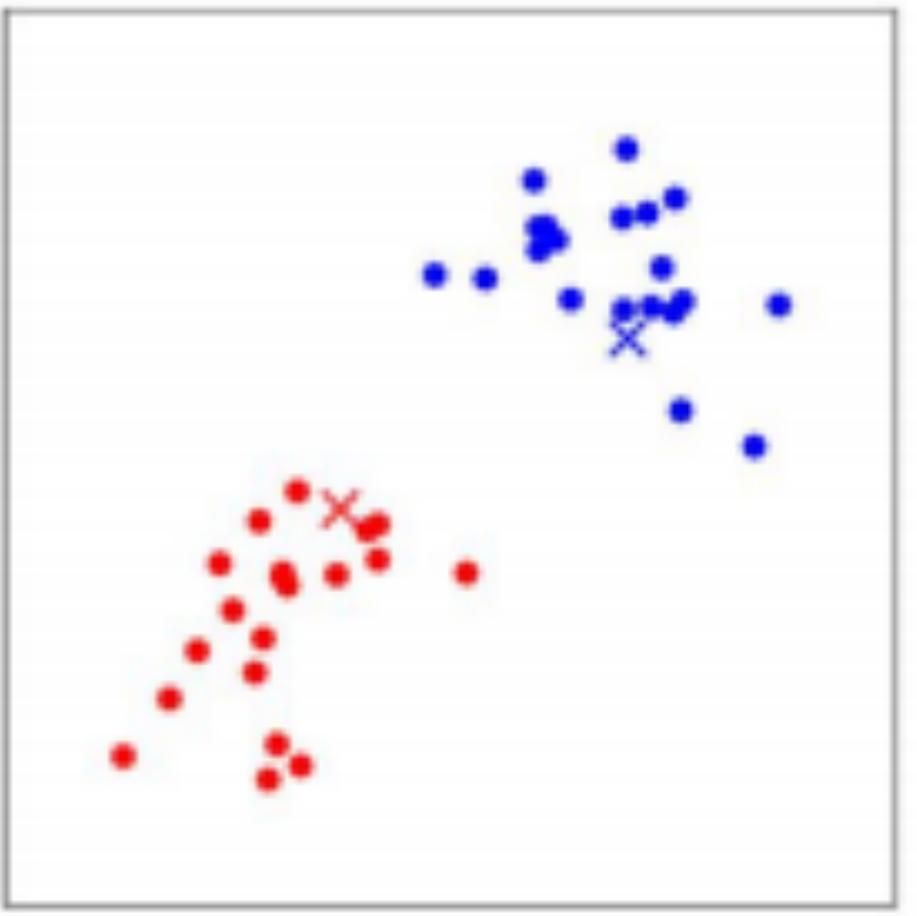
(b)



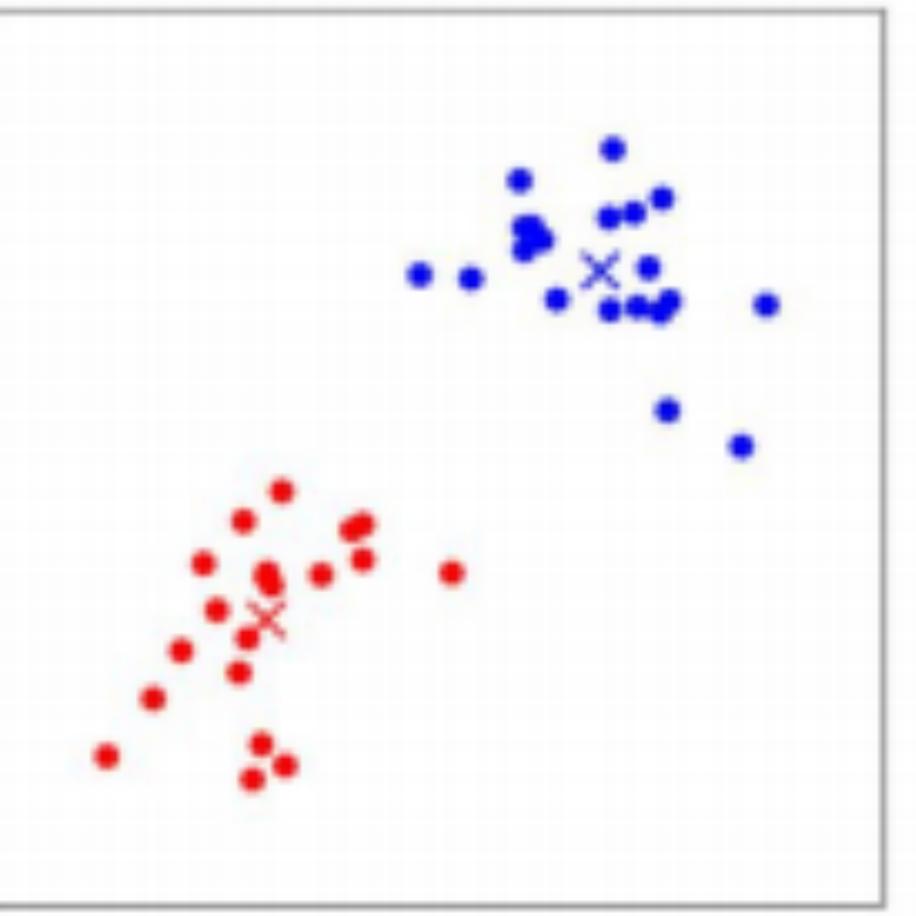
(c)



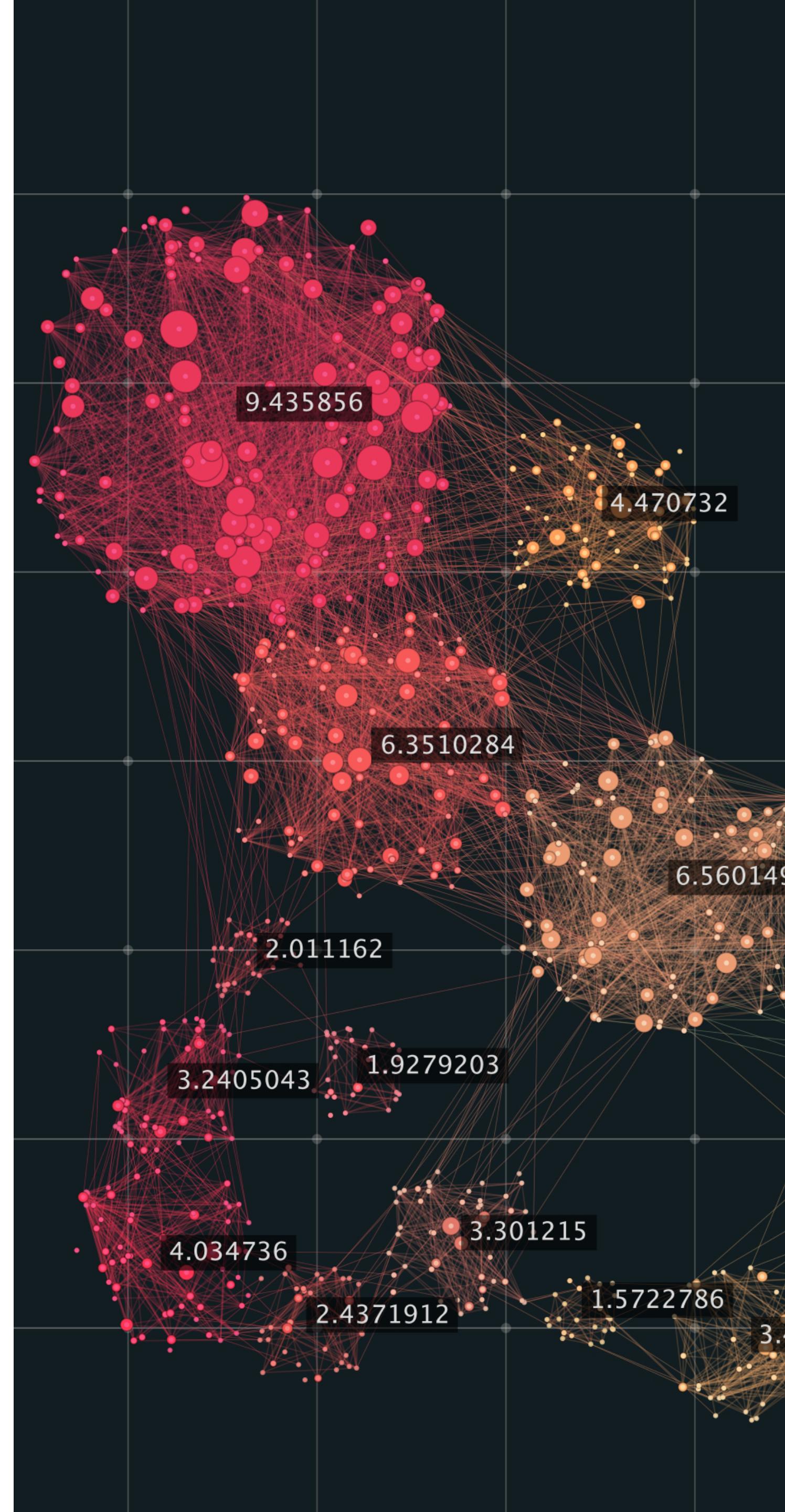
(d)



(e)



(f)



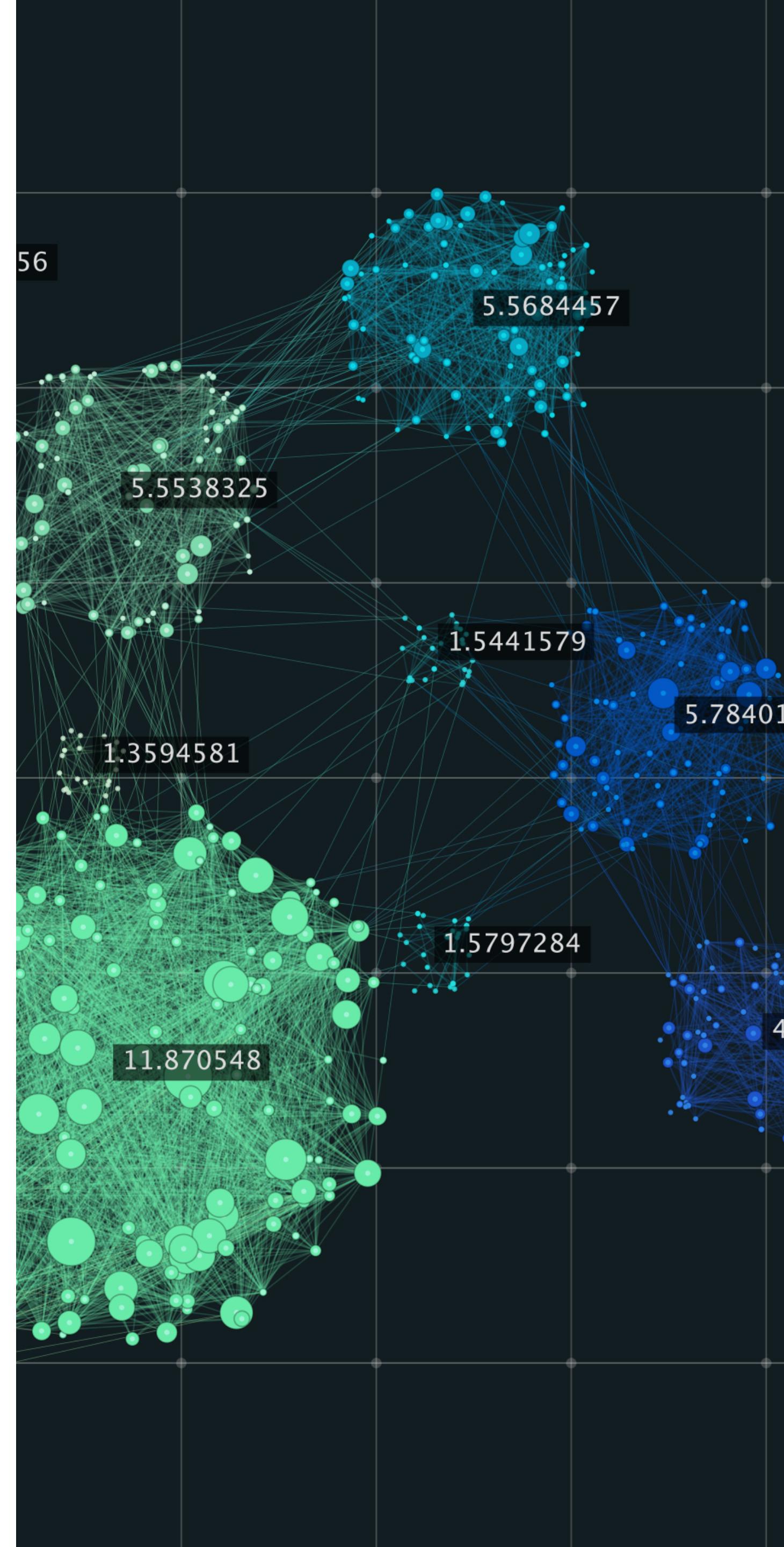
الگوريتم K-means

K-MEANS(P, k)

Input: a dataset of points $P = \{p_1, \dots, p_n\}$, a number of clusters k

Output: centers $\{c_1, \dots, c_k\}$ implicitly dividing P into k clusters

```
1 choose  $k$  initial centers  $C = \{c_1, \dots, c_k\}$ 
2 while stopping criterion has not been met
3   do ▷ assignment step:
4     for  $i = 1, \dots, N$ 
5       do find closest center  $c_k \in C$  to instance  $p_i$ 
6       assign instance  $p_i$  to set  $C_k$ 
7   ▷ update step:
8   for  $i = 1, \dots, k$ 
9     do set  $c_i$  to be the center of mass of all points in  $C_i$ 
```



مراحل انجام آزمایش ۱

- ۱- مقدار k را به تعداد کلاستر مورد نظر انتخاب کنید.
- ۲- k نقطه متفاوت از دادهها را به عنوان اولین نقاط مرکز کلاستر (centroid) انتخاب کنید.

تابعی بنویسید که مراکز کلاسترها را به صورت تصادفی انتخاب کند. ورودی این الگوریتم تعداد کلاس‌ها و خروجی آن مراکز کلاستر باشد.

- ۳- با معیار فاصله، هر یک از دادهها را به یک مرکز کلاستر تخصیص دهید.

تابعی بنویسید که یک داده و مراکز دسته‌ها را به عنوان ورودی بگیرد و اندیس نزدیکترین دسته به داده ورودی در آرایه را برگرداند.

- ۴- برای هر کلاستر، مرکز کلاستر جدید تعریف کنید.

تابعی بنویسید که آرایه ای از داده‌ها از ورودی دریافت کند و سپس میانگین داده‌های ورودی را باز گرداند.

- ۵- میزان تغییرات کلاستر جدید و قدیم را مقایسه کنید. در صورت همگرایی، آموزش پایان می‌پذیرد و در غیر این صورت به گام ۴ برگردید.

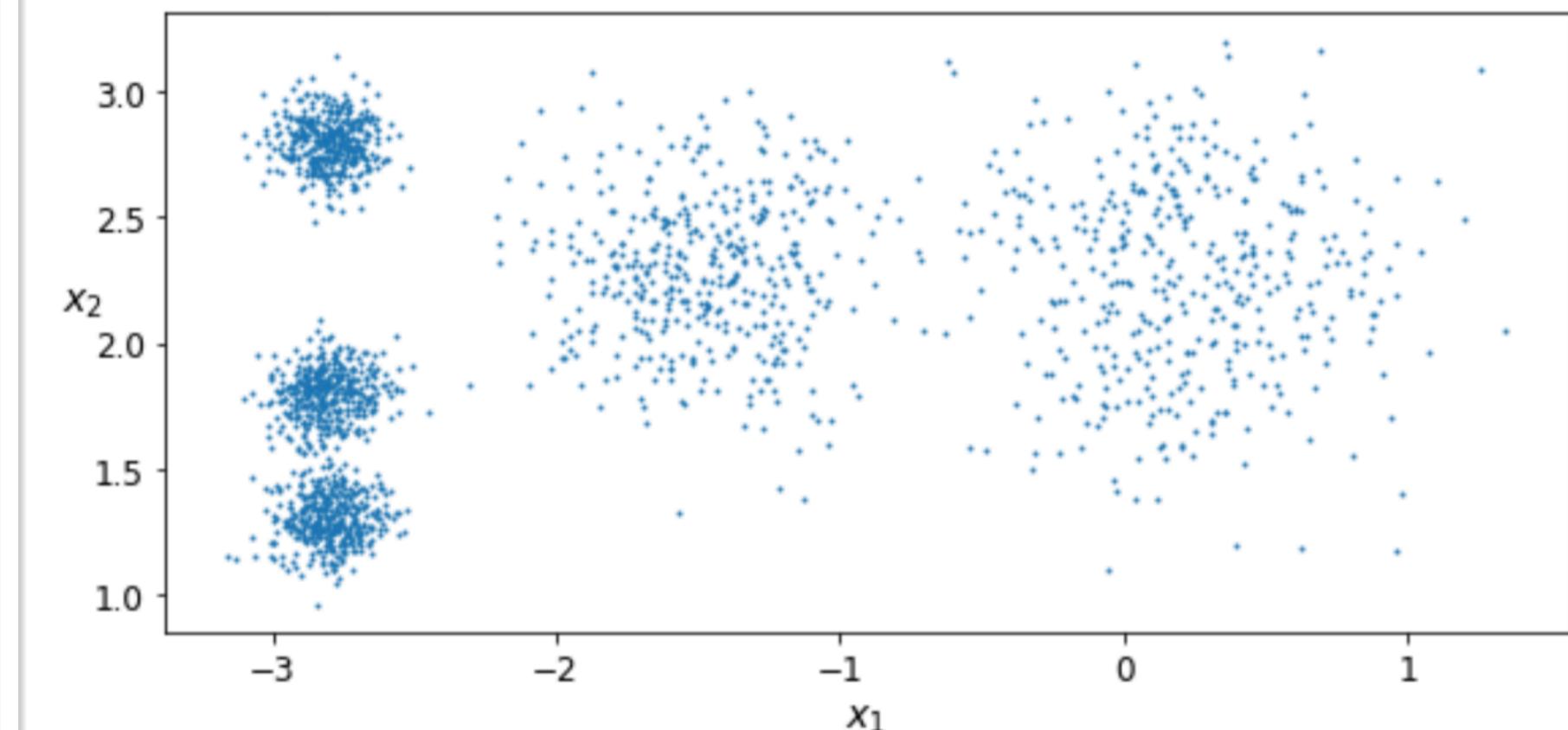
تست الگوريتم

مجموعه داده

```
from sklearn.datasets import make_blobs

blob_centers = np.array(
    [[ 0.2,  2.3],
     [-1.5 ,  2.3],
     [-2.8,  1.8],
     [-2.8,  2.8],
     [-2.8,  1.3]])
blob_std = np.array([0.4, 0.3, 0.1, 0.1, 0.1])

X, y = make_blobs(n_samples=2000, centers=blob_centers,
                  cluster_std=blob_std, random_state=7)
```



ارزیابی الگوریتم

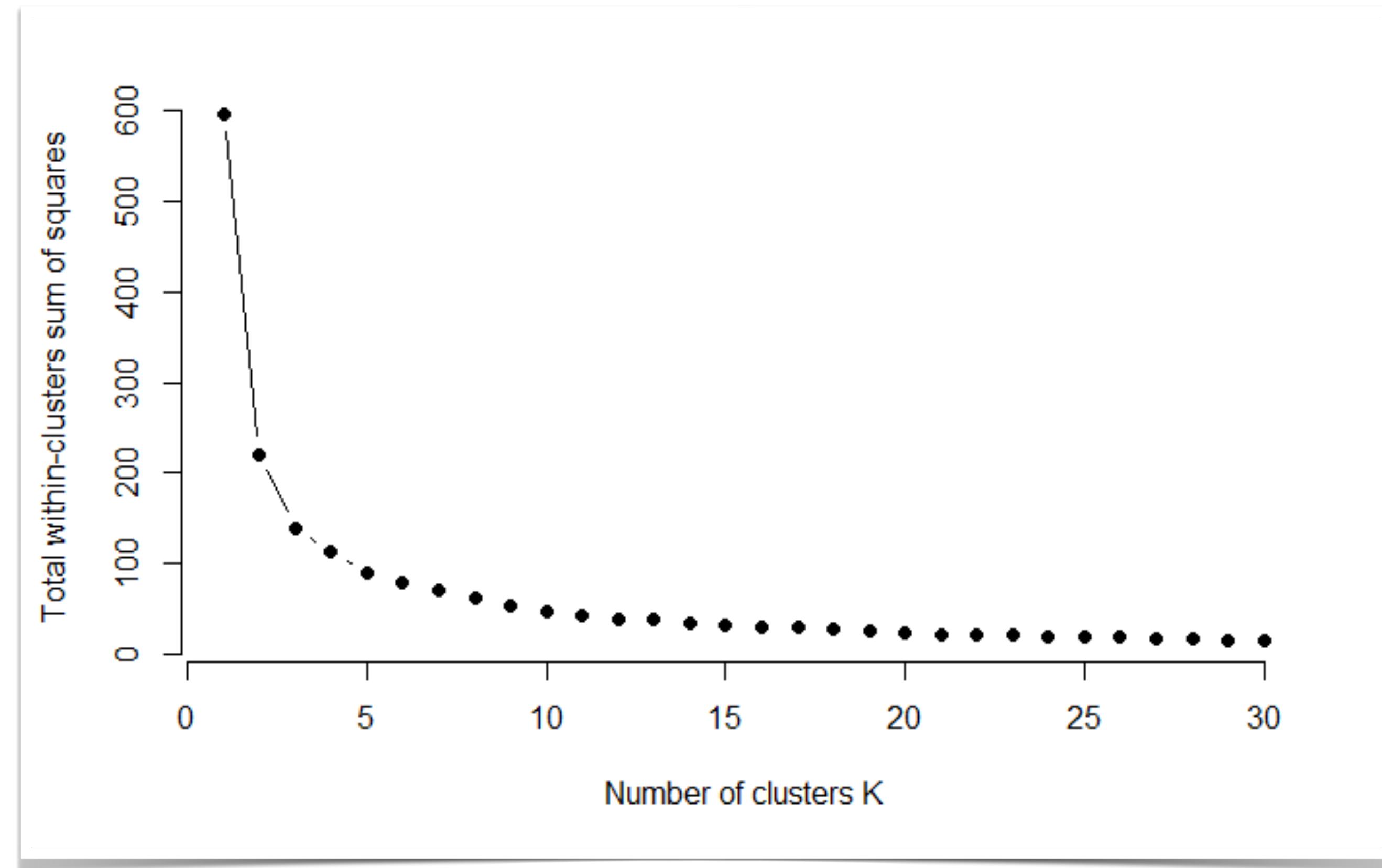
۱- تابعی بنویسید که نقاط و کلاسترها متناظر با آن ها از ورودی بگیرد، سپس مجموع میانگین فاصله نقاط هر کلاستر را تا مرکز کلاستر محاسبه کند.

WCSS (Within Cluster Sum of Squares)=

$$\sum_{C_j} \sum_{P_i \text{ in } C_j} distance(P_i, C_j)^2$$

۲- برای تعداد مختلف کلاستر، WCSS را بیابید و با توجه به این مقدار، تعداد بهینه کلاستر را مشخص کنید.

ارزیابی الگوریتم



مراحل انجام آزمایش ۲

کاهش حجم تصویر به وسیله k-means

- ۱- با استفاده از دستور `imread (filename)` عکس `imagesmall` را از پوشه `Images/Exp04` پوشید.
گیت‌هاب درس بارگذاری کنید.
۲. برای استفاده از الگوریتم k-means، ابتدا آرایه 3×3 بعدی عکس را به یک آرایه 2×2 بعدی تبدیل کنید که در آن هر سطر متناظر با یک پیکسل خواهد بود.
۳. الگوریتم k-means را با تعداد کلاستر برابر با $4, 6, 8, 10$ و 2 اجرا کنید و سپس مقدار هر پیکسل را با مقدار مرکز دسته متناظر جایگذاری کنید.
- ۴- عکس خروجی را رسم کنید و با عکس اولیه مقایسه کنید.

کاهش حجم تصویر به وسیله k-means



K-means محدودیت‌های

