

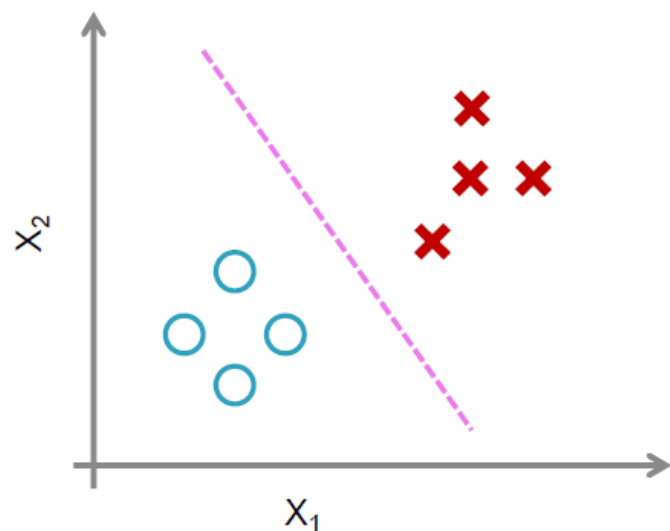


## فصل یازدهم : فهرست مطالب

- یادگیری بدون نظارت
- کاربردها
- خوشه‌بندی
- الگوریتم K-means
- بهبود خوشه‌بندی
- الگوریتم دو بخشی‌ساز
- خوشه‌بندی سلسله‌مراتبی

## فصل یازدهم: یادگیری نظارت شده

□ یادگیری نظارت شده. به ازای هر نمونه، پاسخ درست داده شده است.



مجموعه‌ی آموزشی

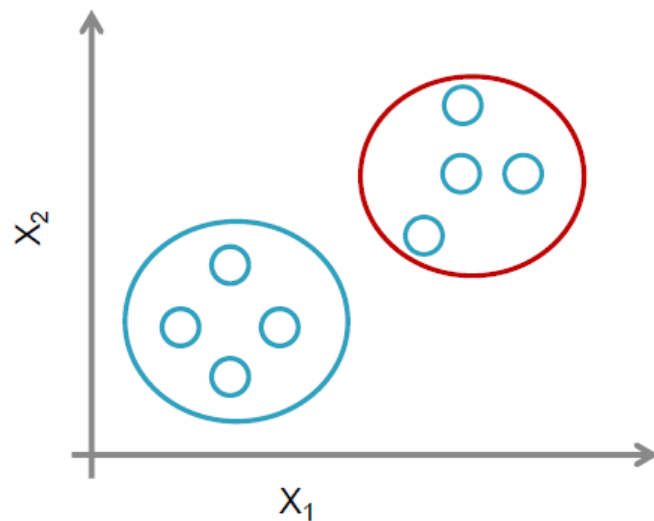
$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), (x^{(3)}, y^{(3)}), \dots, (x^{(m)}, y^{(m)})\}$$

## فصل یازدهم : یادگیری بدون نظارت

□ یادگیری بدون نظارت. عدم آگاهی از پاسخهای درست.

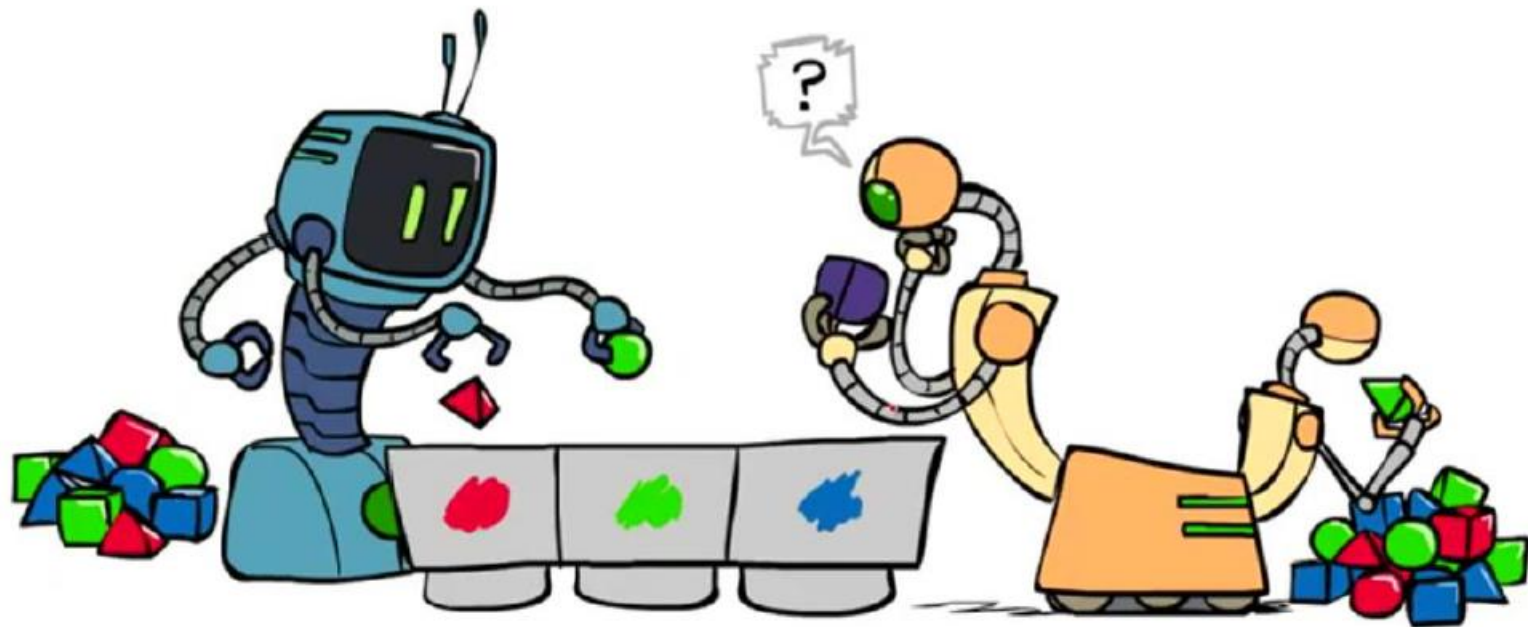
$$\{x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(m)}\}$$

مجموعه‌ی آموزشی



□ هدف. تشخیص ساختار در داده‌های ورودی [گروه‌بندی داده‌های مشابه]

## فصل یازدهم : خوشه بندی



## فصل یازدهم: کاربرد خوشه بندی (گروه بندی اخبار مرتبط)

Google News

news.google.com

Web Images Videos Maps News Shopping Gmail more

andrewyantakng@gmail.com | Web History | Settings | Sign out

Google news

Search News Search the Web

Advanced news search

U.S. edition Add a section

### Top Stories

Deepwater Horizon  
Fed meeting  
Foreign exchange market  
Lindsay Lohan  
IBM  
Tom Brady  
Toronto  
International Film Festival  
Paris Hilton  
Iran  
Hurricane Igor

### Starred

San Francisco Bay Area  
World  
U.S.  
Business  
Sci/Tech  
More Top Stories  
Spotlight  
Health  
Sports  
Entertainment

### All news

Headlines  
Images

### Top Stories

Christine O'Donnell »  
**White House official denies Tea Party-focused ad campaign**  
CNN International - Ed Henry - 1 hour ago  
Democratic sources say the White House is not considering an ad campaign tying Republicans to the Tea Party. Washington (CNN) -- A top White House official sharply denied a report that claims President Obama's political advisers are weighing a national ...  
Tea Party is misplacing the blame, former President Bill Clinton claims  
New York Daily News  
GOP tea party backer defends Christine O'Donnell The Associated Press  
Atlanta Journal Constitution - Politics Daily - MyFox Washington DC - Salon  
all 726 news articles »

**US Stocks Climb After Recession Called Over, Homebuilders Gain**  
MarketWatch - Kristina Peterson - 16 minutes ago  
NEW YORK (MarketWatch) -- US stocks climbed Monday, gaining speed after a key nonprofit organization officially called the recession over, giving investors a boost of confidence in the gradual economic recovery.  
Longest recession since 1930s ended in June 2009, group says  
Los Angeles Times  
Downturn Was Longest in Decades, Panel Confirms New York Times  
Wall Street Journal - AFP - CNN - USA Today  
all 276 news articles »

**BP Oil Well, Site of National Catastrophe, Dies at One**  
Vanity Fair - Juli Weiner - 22 minutes ago  
The BP oil well, site of the Deepwater Horizon explosion that led to the worst oil spill in US history, died today at one year old.  
Video: Blown-out BP Well Finally Killed in Gulf The Associated Press  
Weiss Doubts BP Would End Operations in Gulf of Mexico: Video Bloomberg  
CNN International - Wall Street Journal (blog) - The Guardian - New York Times  
all 2,292 news articles »

CNN Intern...

RECESSION

MyFox Phila...

Reuters

### Recent

Recession officially ended in June 2009  
CNNMoney - Chris Isidore - 39 minutes ago  
Hurricane Igor lashes Bermuda  
USA Today - Gerry Broome - 5 minutes ago  
Explain what you want from us: reads front-page editorial  
msnbc.com - Olivia Torres - 10 minutes ago

Crisis response: Pakistan floods

San Francisco Bay Area - Edit

Clorox »  
Bay Biz Buzz: Clorox close to selling STP, Armor All  
San Jose Mercury News - 48 minutes ago - all 24 articles »

Google's official beekeeper keeps the company buzzing with excitement  
San Jose Mercury News - Bruce Newman - 1 hour ago

Jon Sylva »  
Martinez man still unconscious as police investigate weekend shooting  
San Jose Mercury News - Robert Salonga - 48 minutes ago - all 6 articles »

### Spotlight

Sarkozy rages at EU 'humiliation'  
Financial Times - Peggy Hollinger - Sep 16, 2010

# فصل یازدهم: کاربرد خوشه بندی (گروه بندی اخبار مرتبط)



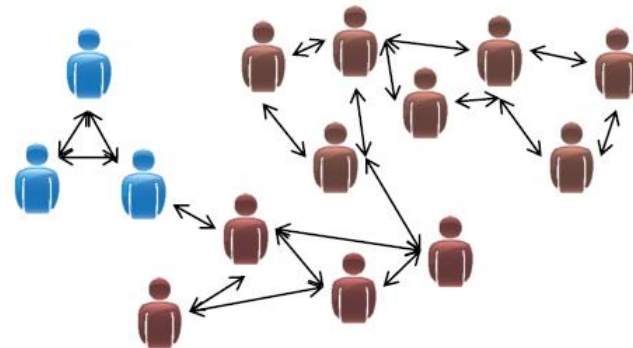
## فصل یازدهم: چند کاربرد دیگر از یادگیری بدون نظارت



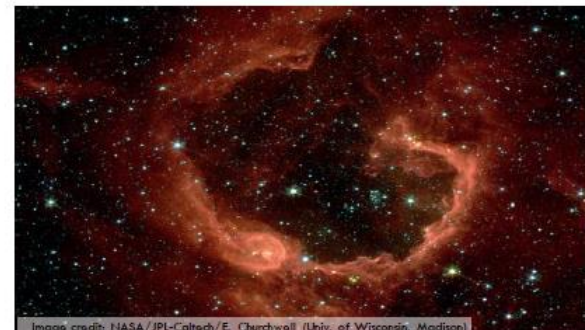
سازماندهی کلاسترهای مناسباتی (مرکز داده‌ها)



بخش‌بندی بازار



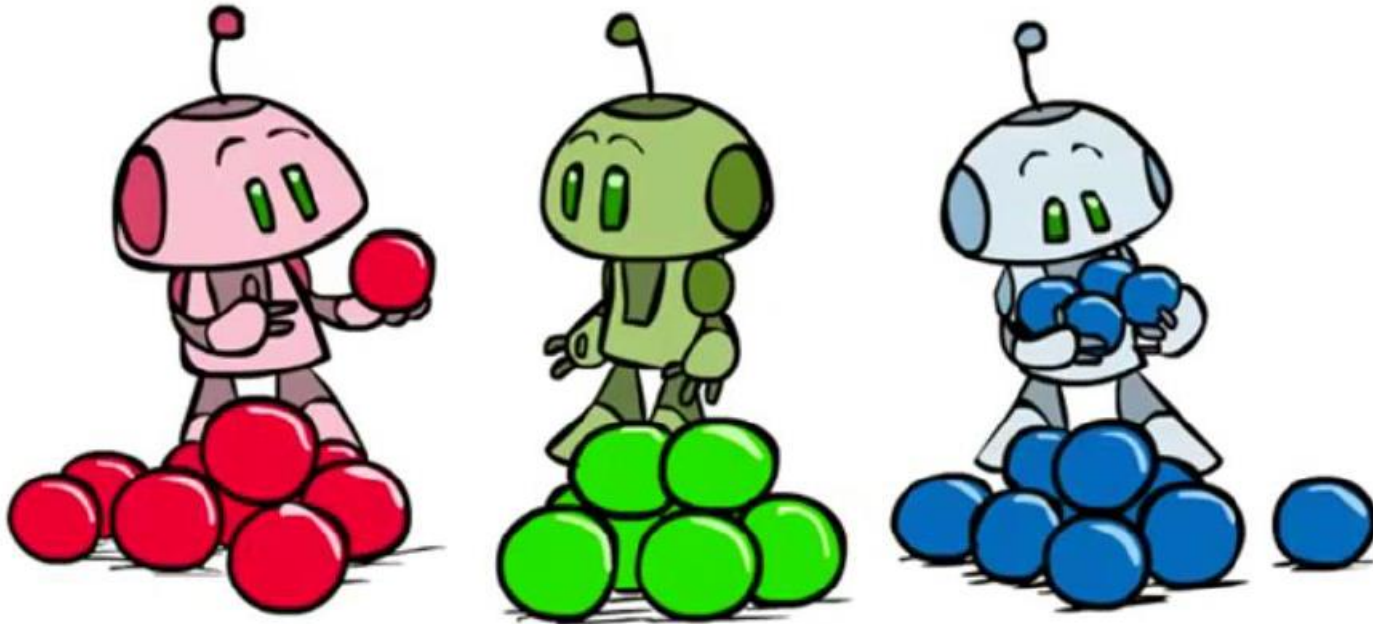
تحلیل شبکه‌های اجتماعی



تحلیل داده‌های ستاره‌شناسی (نوعه تشکیل کهکشان‌ها)



## فصل یازدهم: الگوریتم خوشه بندی K-means



## فصل یازدهم: الگوریتم خوشه بندی K-means

□ یک الگوریتم خوشه بندی تکرار شونده.

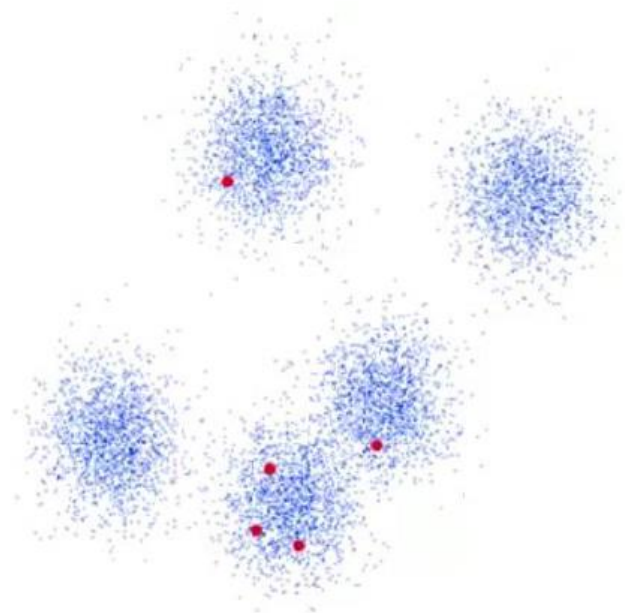
□ K نقطه را به صورت تصادفی به عنوان مراکز خوشه ها انتخاب کن.

□ مراحل زیر را تکرار کن:

■ هر داده را به یک خوشه با نزدیک ترین مرکز انتساب بده.

■ مرکز هر خوشه را با میانگین گیری از داده های انتساب یافته به آن خوشه به روز رسانی کن.

□ توقف: زمانی که در یک تکرار هیچ داده ای خوشه ی خود را عوض نکند.



## فصل یازدهم: الگوریتم خوشه بندی K-means

□ یک الگوریتم خوشه بندی تکرار شونده.

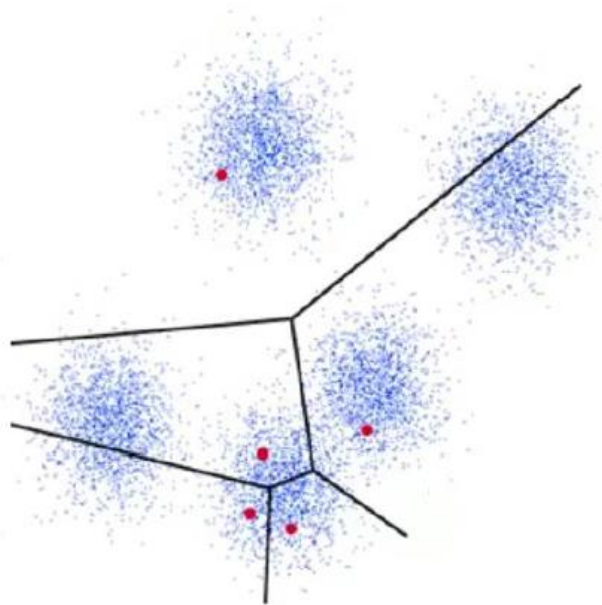
□ K نقطه را به صورت تصادفی به عنوان مراکز خوشه ها انتخاب کن.

□ مراحل زیر را تکرار کن:

■ هر داده را به یک خوشه با نزدیک ترین مرکز انتساب بده.

■ مرکز هر خوشه را با میانگین گیری از داده های انتساب یافته به آن خوشه به روز رسانی کن.

□ توقف: زمانی که در یک تکرار هیچ داده ای خوشه ی خود را عوض نکند.



## فصل یازدهم: الگوریتم خوشه بندی K-means

□ یک الگوریتم خوشه بندی تکرار شونده.

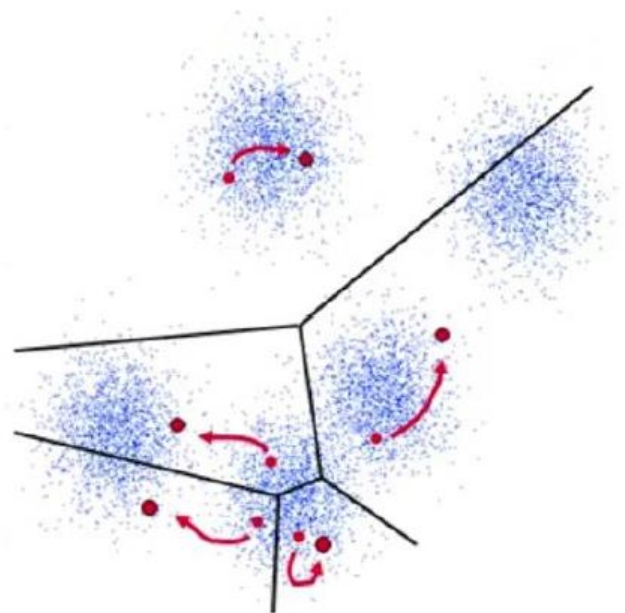
□ K نقطه را به صورت تصادفی به عنوان مراکز خوشه ها انتخاب کن.

□ مراحل زیر را تکرار کن:

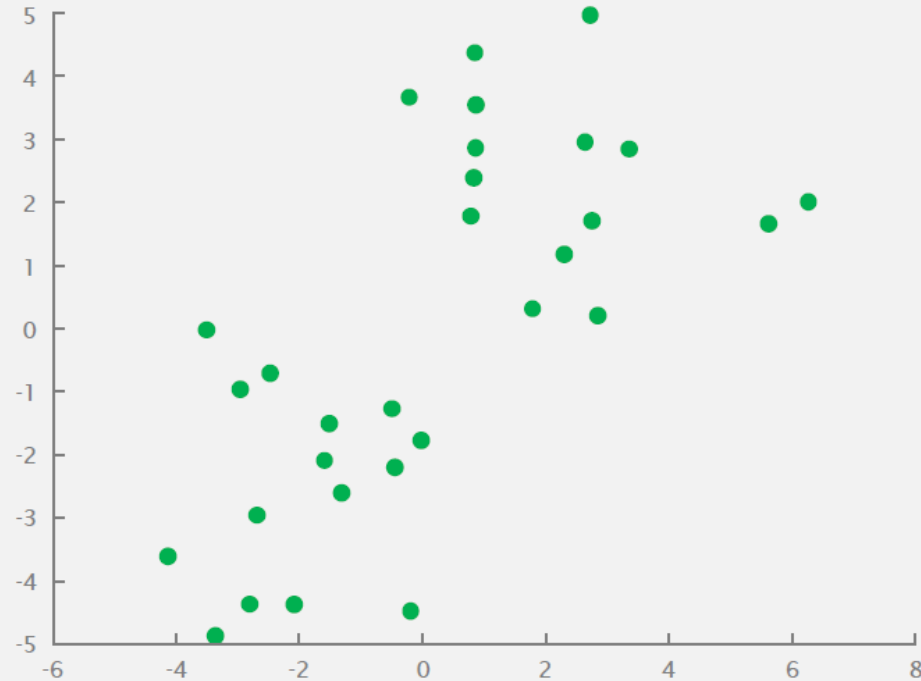
■ هر داده را به یک خوشه با نزدیک ترین مرکز انتساب بده.

■ مرکز هر خوشه را با میانگین گیری از داده های انتساب یافته به آن خوشه به روز رسانی کن.

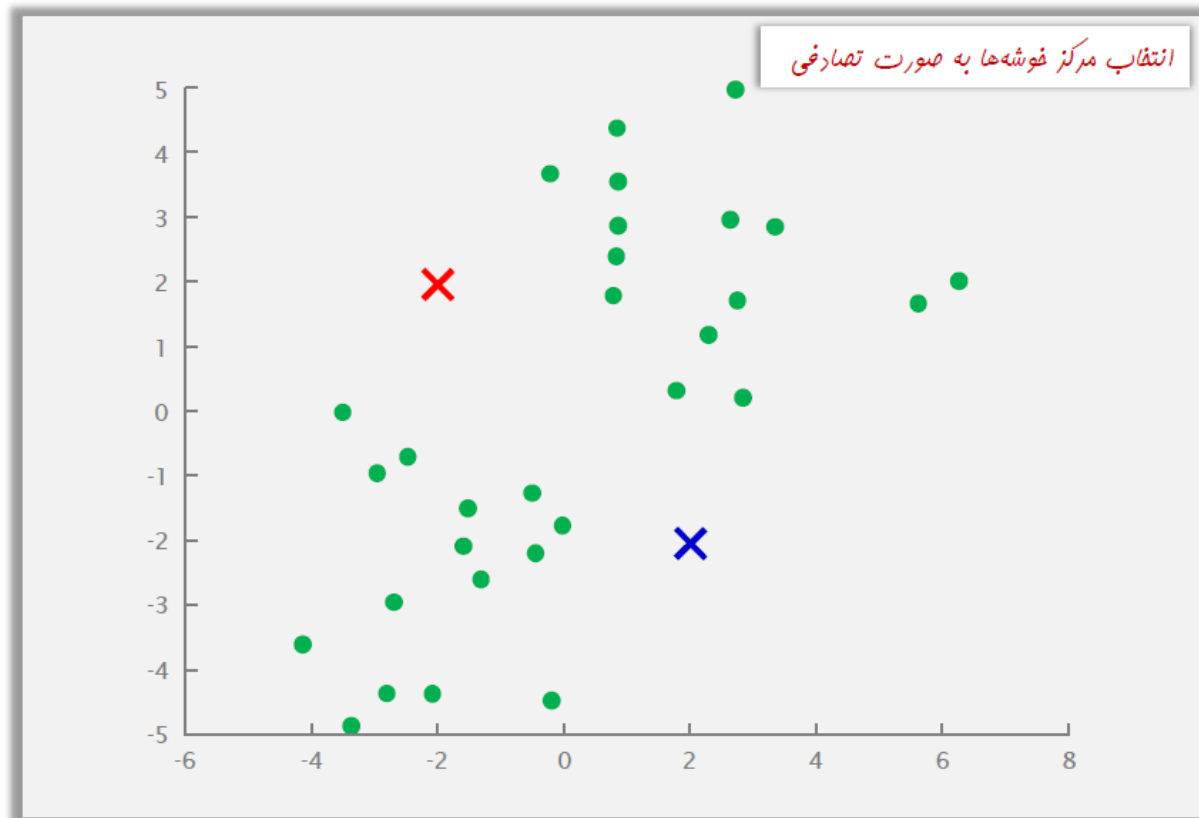
□ توقف: زمانی که در یک تکرار هیچ داده ای خوشه ی خود را عوض نکند.



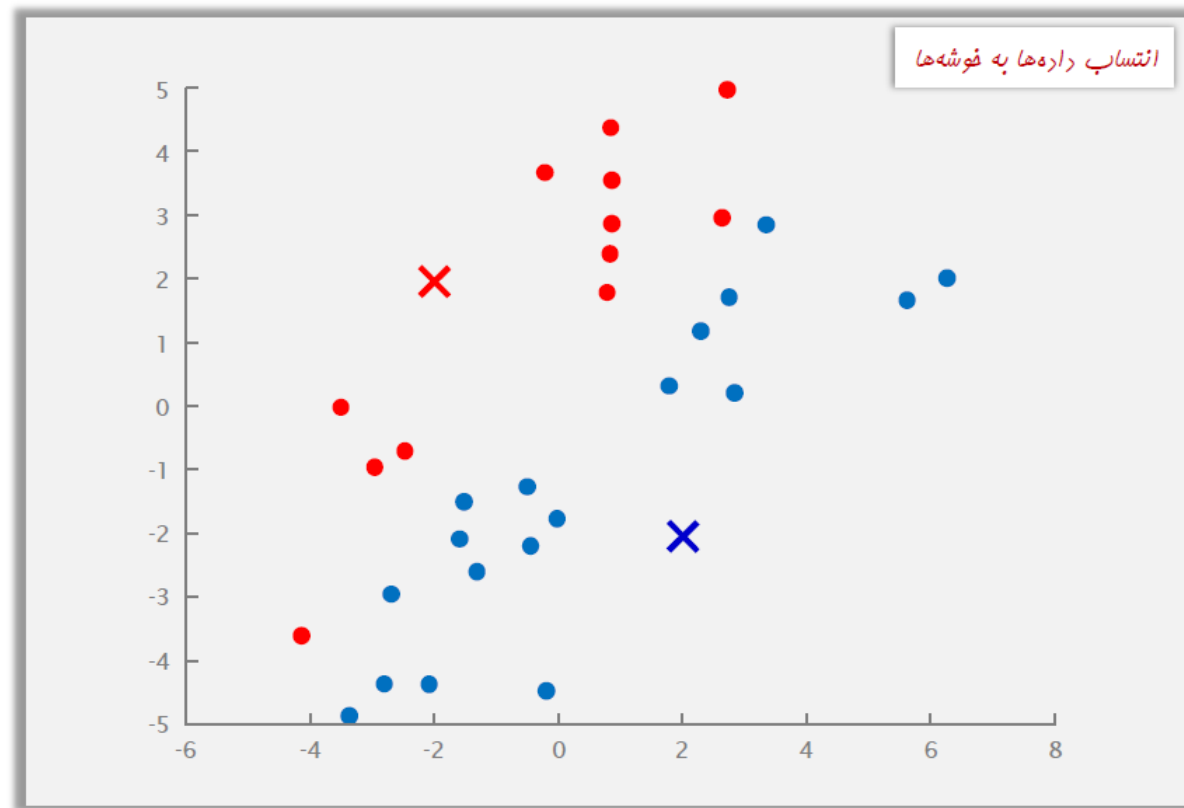
## فصل یازدهم : الگوریتم K-means (اجرای نمایش)



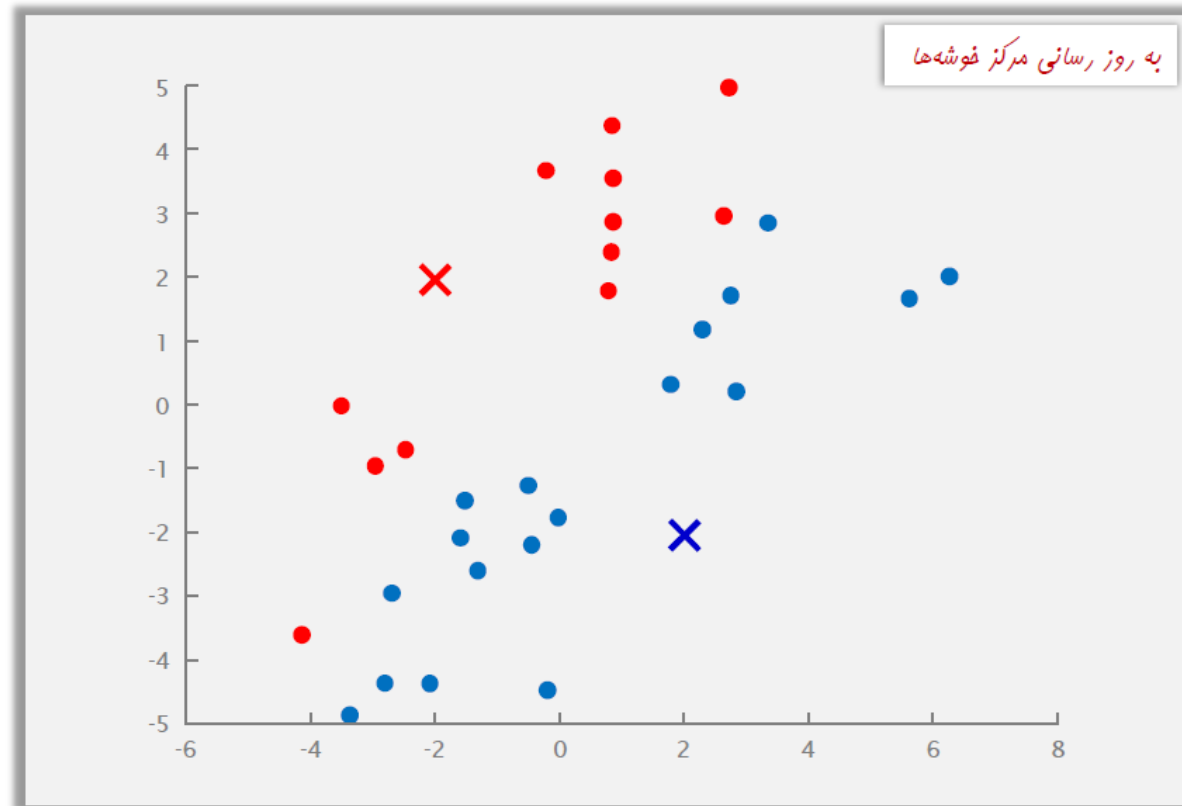
## فصل یازدهم: الگوریتم K-means (اجرای نمایش)



## فصل یازدهم: الگوریتم K-means (اجرای نمایش)

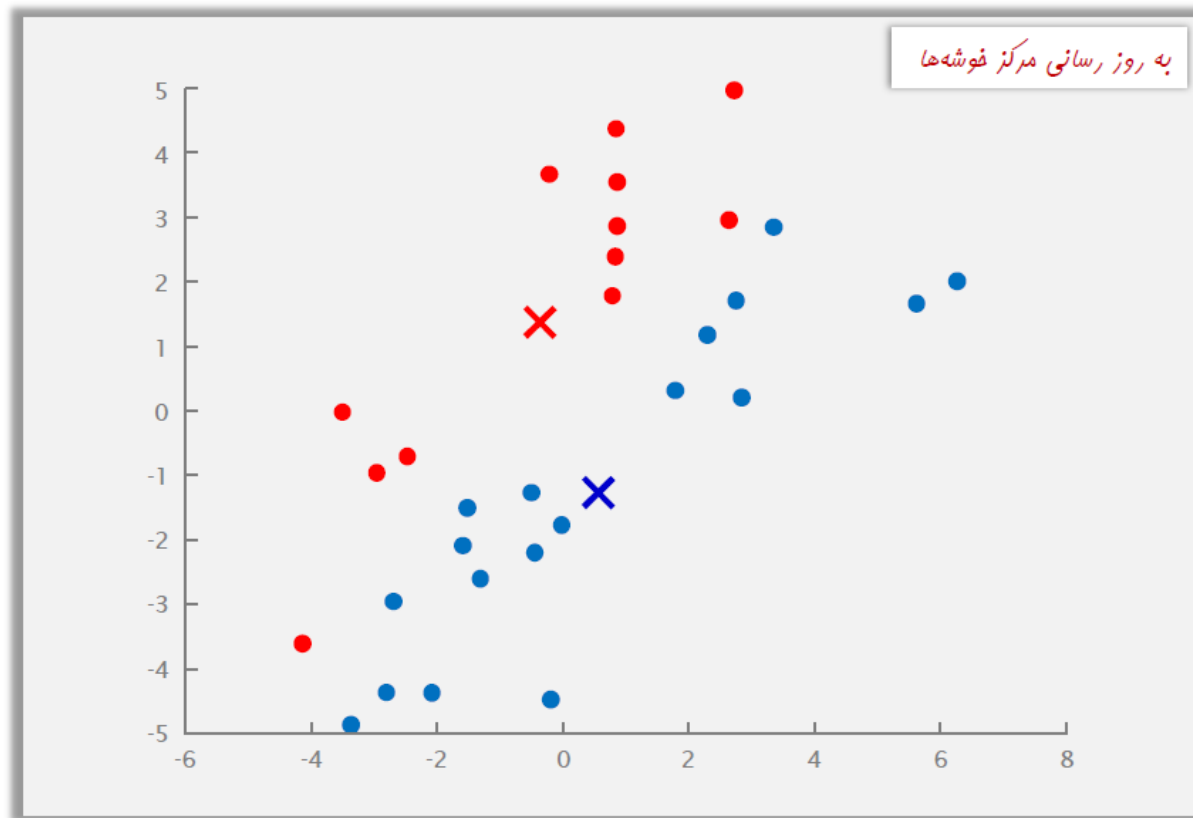


## فصل یازدهم: الگوریتم K-means (اجرای نمایش)

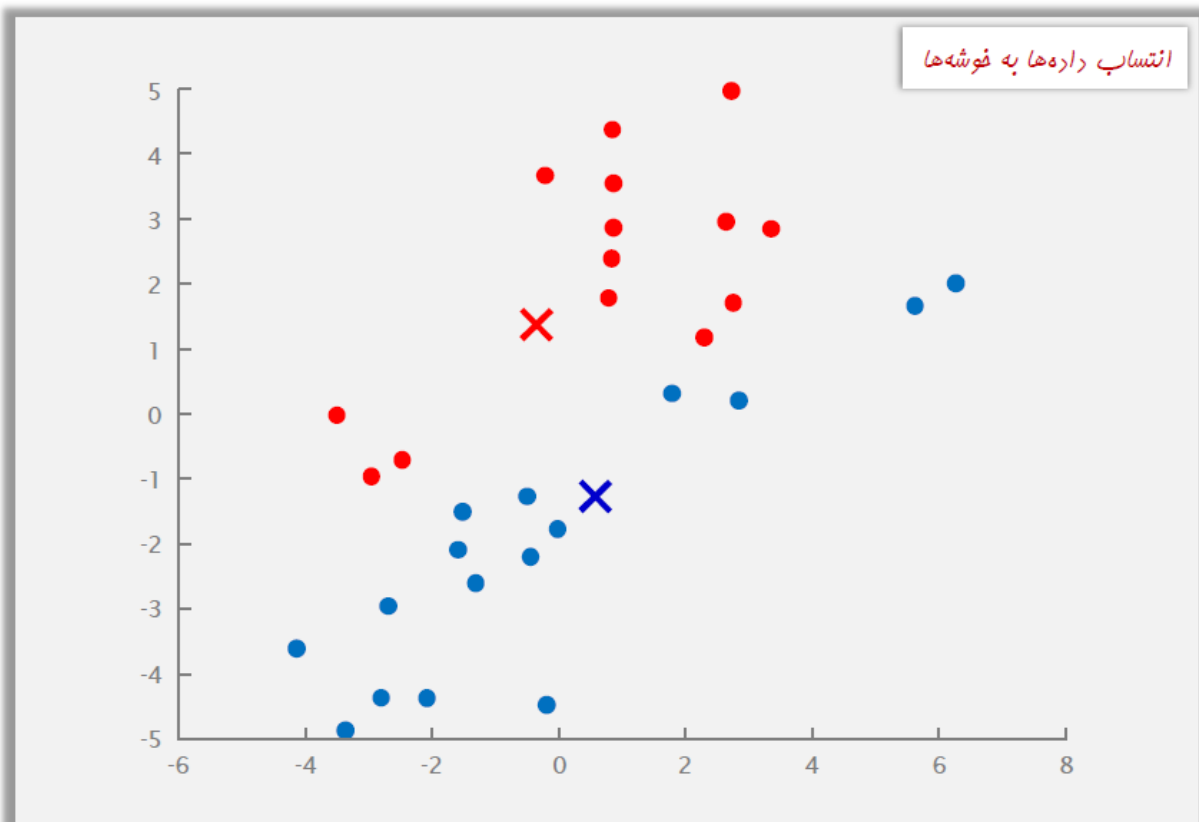




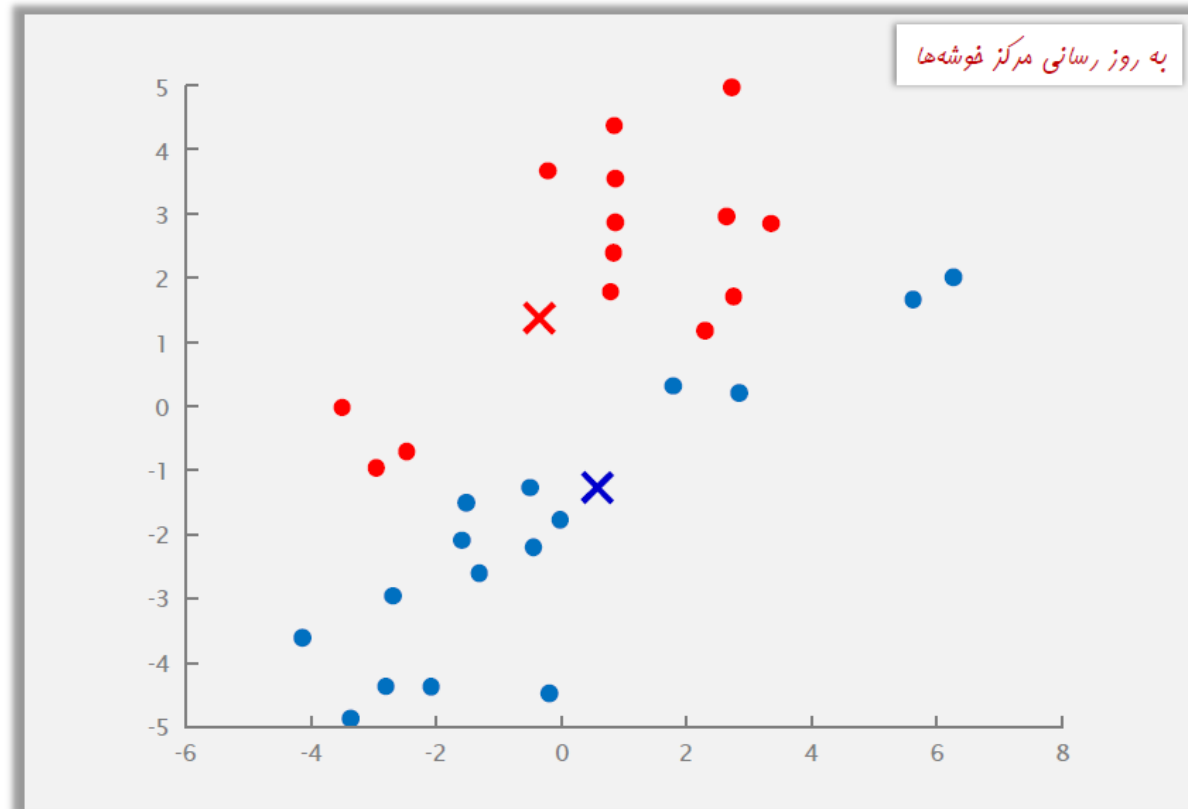
## فصل یازدهم: الگوریتم K-means (اجرای نمایش)



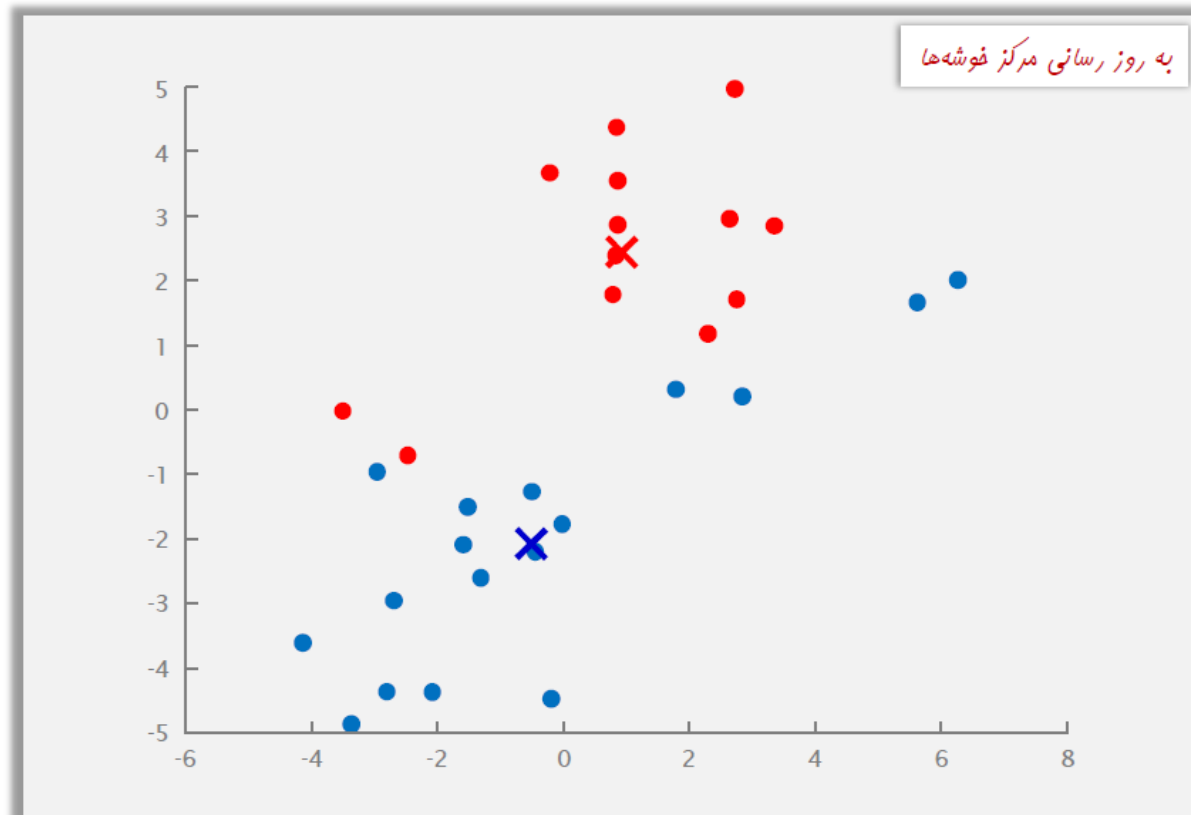
## فصل یازدهم : الگوریتم K-means (اجرای نمایش)



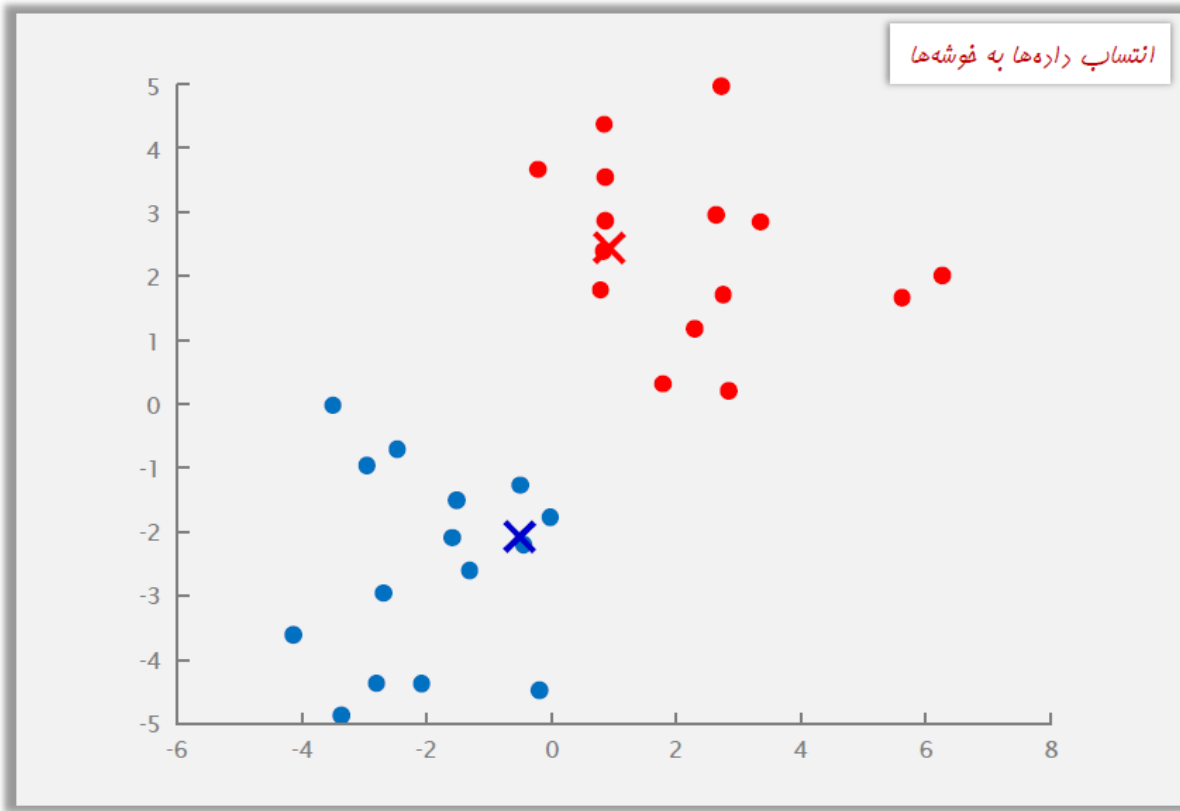
## فصل یازدهم : الگوریتم K-means (اجرای نمایش)



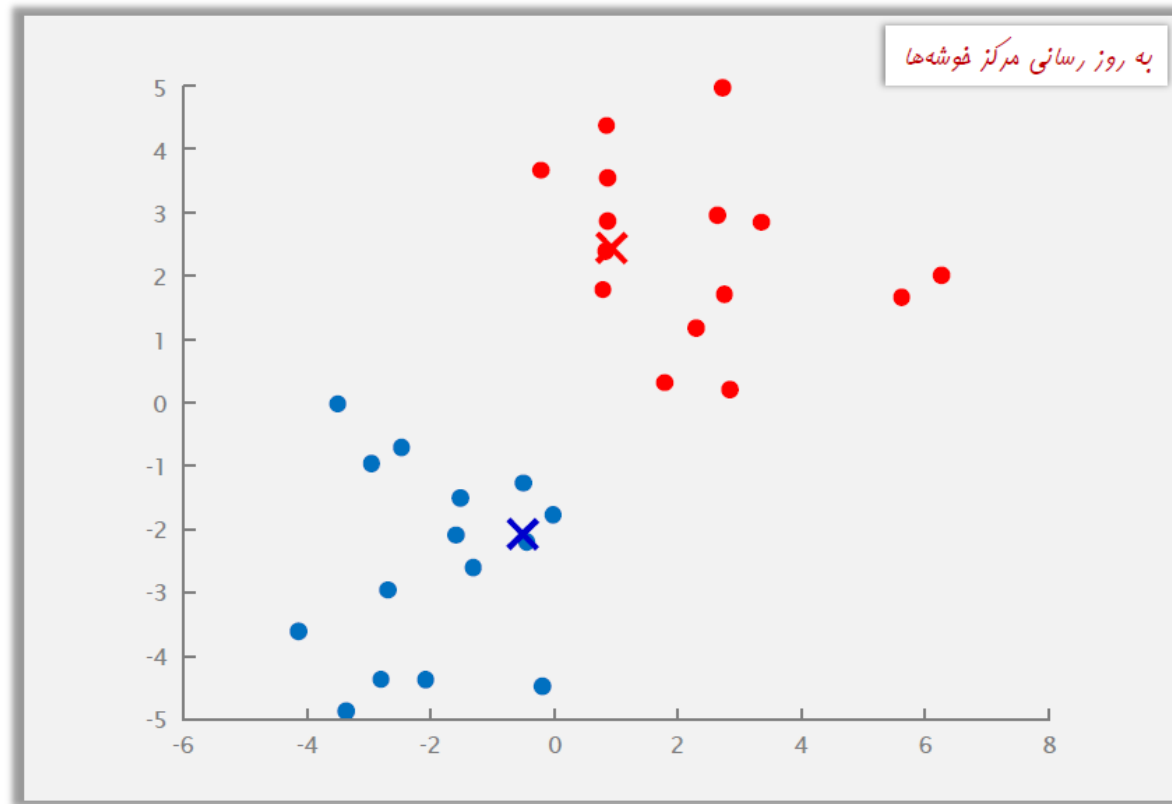
## فصل یازدهم : الگوریتم K-means (اجرای نمایش)



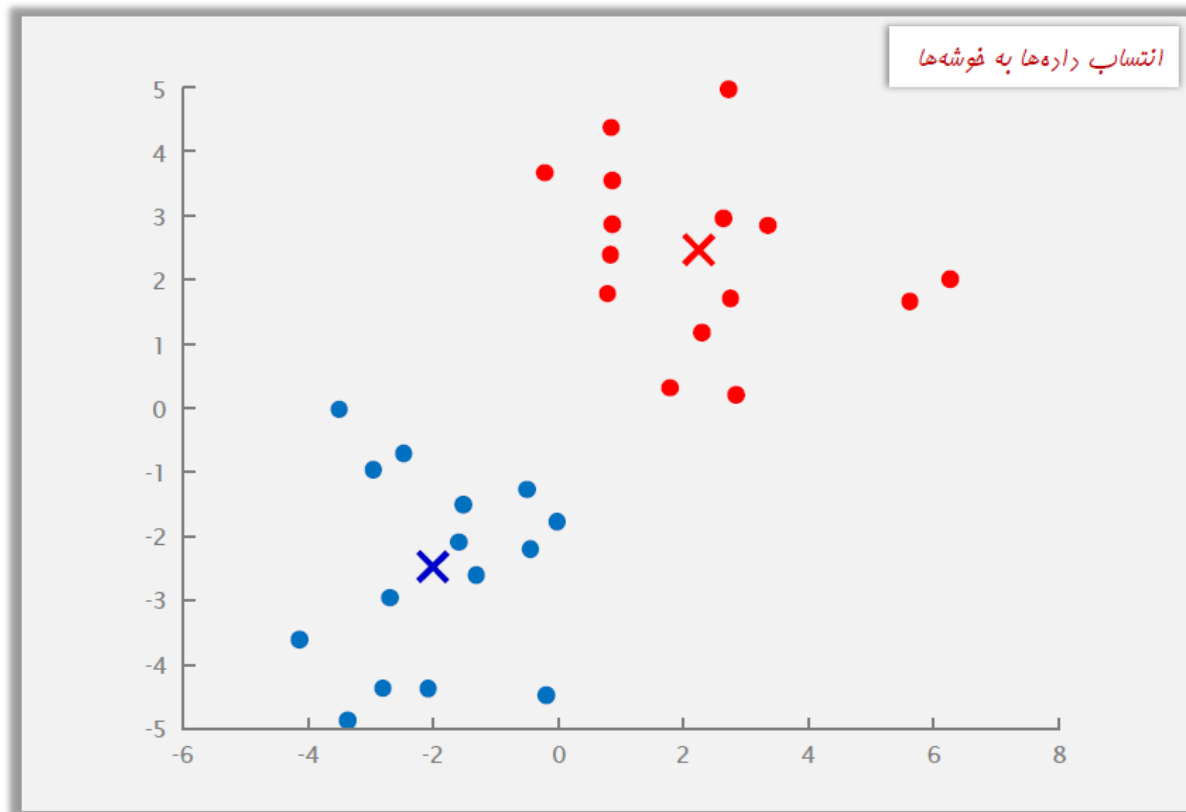
## فصل یازدهم: الگوریتم K-means (اجرای نمایش)



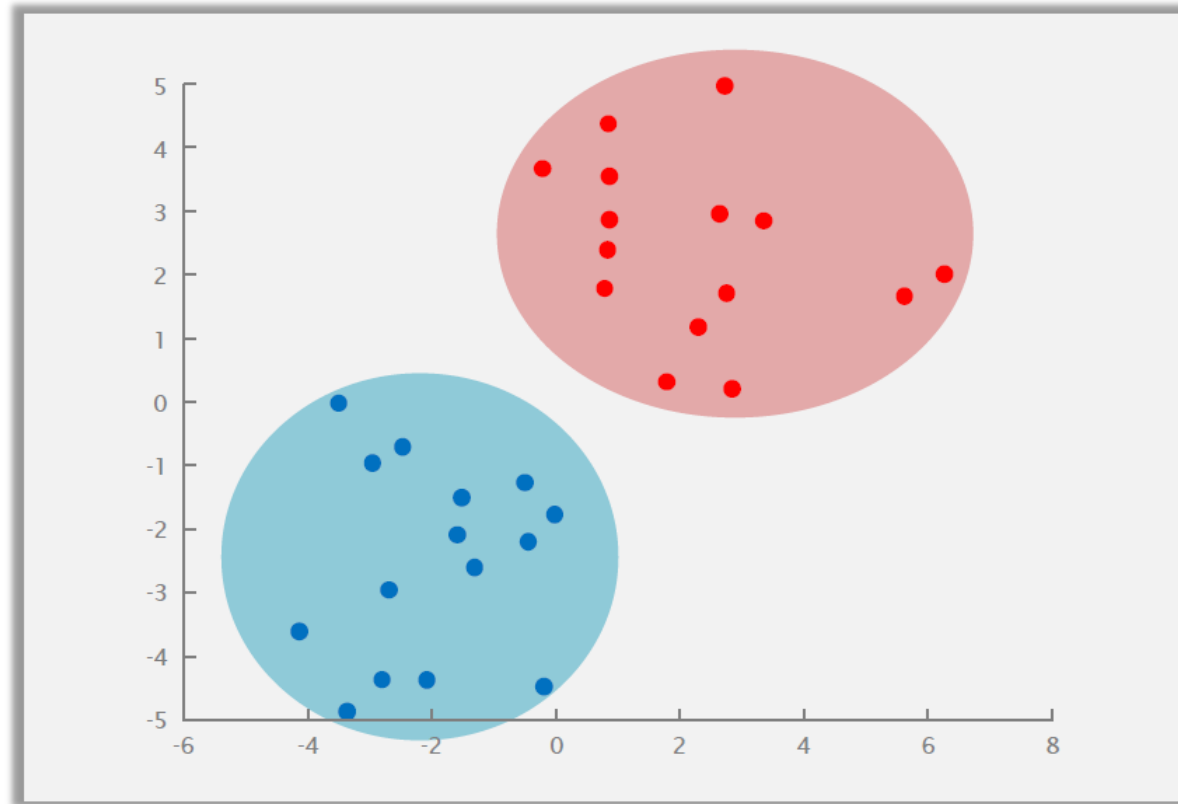
## فصل یازدهم: الگوریتم K-means (اجرای نمایش)



## فصل یازدهم: الگوریتم K-means (اجرای نمایش)



## فصل یازدهم: الگوریتم K-means (اجرای نمایش)





## فصل یازدهم: الگوریتم K-means

□ ورودی.

□ تعداد خوشه‌ها:  $k$

□ مجموعه‌ی آموزشی:  $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$

□ توجه. در مجموعه‌ی آموزشی هیچ برجستگی برای داده‌ها تعیین نشده است.

□ توجه. در خوشه‌بندی نیازی به افزودن ویژگی  $x_0 = 1$  نیست.

## فصل یازدهم : الگوریتم K-means

randomly initialize  $K$  cluster centroids  $\mu_1, \mu_2, \dots, \mu_K \in \mathbb{R}^n$

**repeat**

{

**for**  $i = 1$  **to**  $m$

انتساب داده‌ها به خوشه‌ها

$$c^{(i)} = \arg \min_k \|x^{(i)} - \mu_k\|$$

**for**  $k = 1$  **to**  $K$

به روز رسانی مرکز خوشه‌ها

$\mu_k =$  average of points assigned to cluster  $k$

}

خوشه‌بندی: تابع هدف

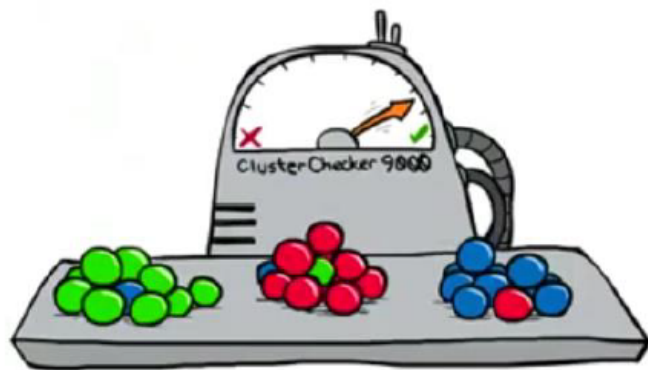
## فصل یازدهم : تابع هدف

□ نمادها.

□  $\mu_k$  : مرکز خوشه‌ی  $k$

□  $c^{(i)}$  : شماره‌ی خوشه‌ی اختصاص یافته به داده‌ی  $x^{(i)}$

□  $\mu_{c^{(i)}}$  : مرکز خوشه‌ی اختصاص یافته به داده‌ی  $x^{(i)}$



□ تابع هدف.

$$J(c^{(1)}, c^{(2)}, \dots, c^{(m)}, \mu_1, \mu_2, \dots, \mu_k) = \frac{1}{m} \sum_{i=1}^m \|x^{(i)} - \mu_{c^{(i)}}\|^2$$

## فصل یازدهم : الگوریتم K-means

randomly initialize  $K$  cluster centroids  $\mu_1, \mu_2, \dots, \mu_K \in \mathbb{R}^n$

**repeat**

{

**for**  $i = 1$  **to**  $m$

$$c^{(i)} = \arg \min_k \|x^{(i)} - \mu_k\|$$

کمینه سازی تابع هدف  
نسبت به پارامترهای  $c^{(i)}$

**for**  $k = 1$  **to**  $K$

$\mu_k =$  average of points assigned to cluster  $k$

کمینه سازی تابع هدف  
نسبت به پارامترهای  $\mu$

}

## مقداردهی اولیه به مراکز خوشه‌ها

## فصل یازدهم : الگوریتم K-means

randomly initialize  $K$  cluster centroids  $\mu_1, \mu_2, \dots, \mu_k \in \mathbb{R}^n$

**repeat**

{

**for**  $i = 1$  **to**  $m$

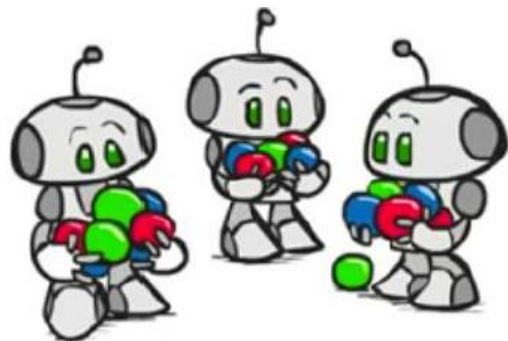
$$c^{(i)} = \arg \min_k \|x^{(i)} - \mu_k\|$$

**for**  $k = 1$  **to**  $K$

$\mu_k =$  average of points assigned to cluster  $k$

}

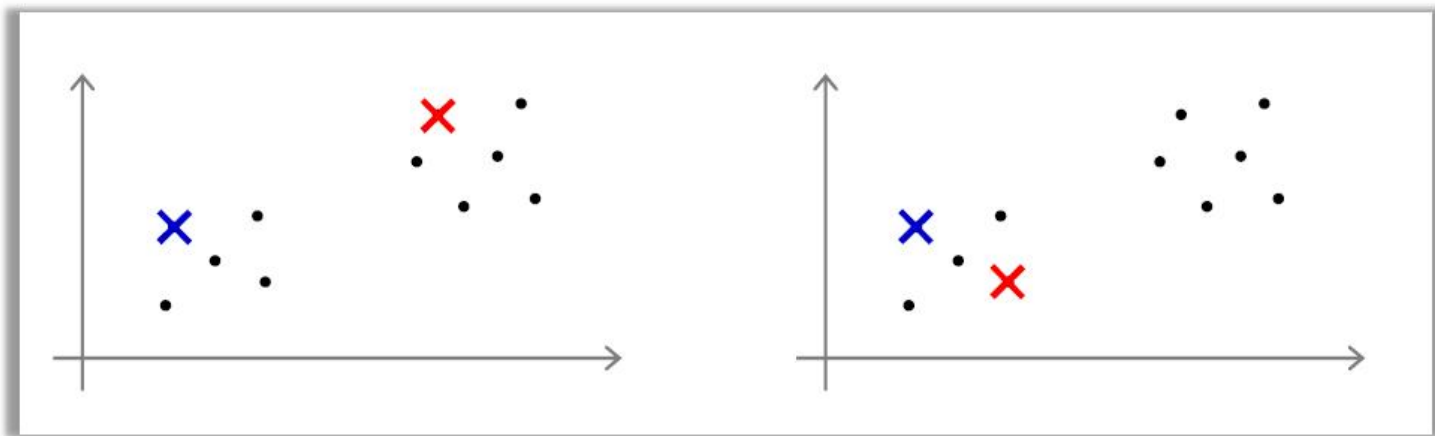
## فصل یازدهم : مقداردهی اولیه به مرکز خوشه ها



□ مقداردهی اولیه.  $[K \leq m]$

□ انتخاب  $K$  نمونه‌ی آموزشی به صورت تصادفی

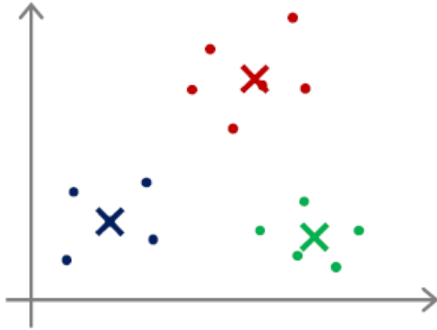
□ انتساب مراکز خوشه‌ها به  $K$  نمونه‌ی انتخاب شده



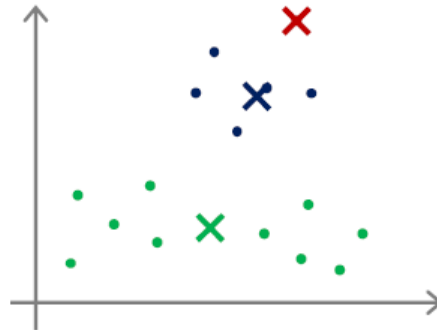
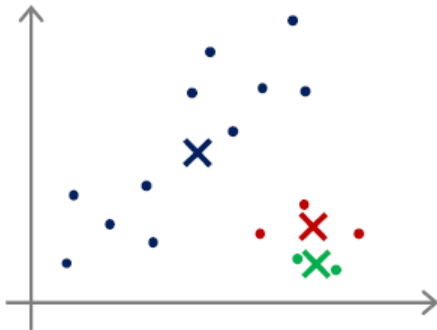


## فصل یازدهم : بهینه ی محلی

□ بهینه ی سراسری.



□ بهینه ی محلی.



## فصل یازدهم: اجتناب از بهینه های محلی

**for**  $t = 1$  **to**  $MAX$

{

randomly initialize cluster centroids  $\mu_1, \mu_2, \dots, \mu_k$

run K-means to get  $c^{(1)}, c^{(2)}, \dots, c^{(m)}, \mu_1, \mu_2, \dots, \mu_k$

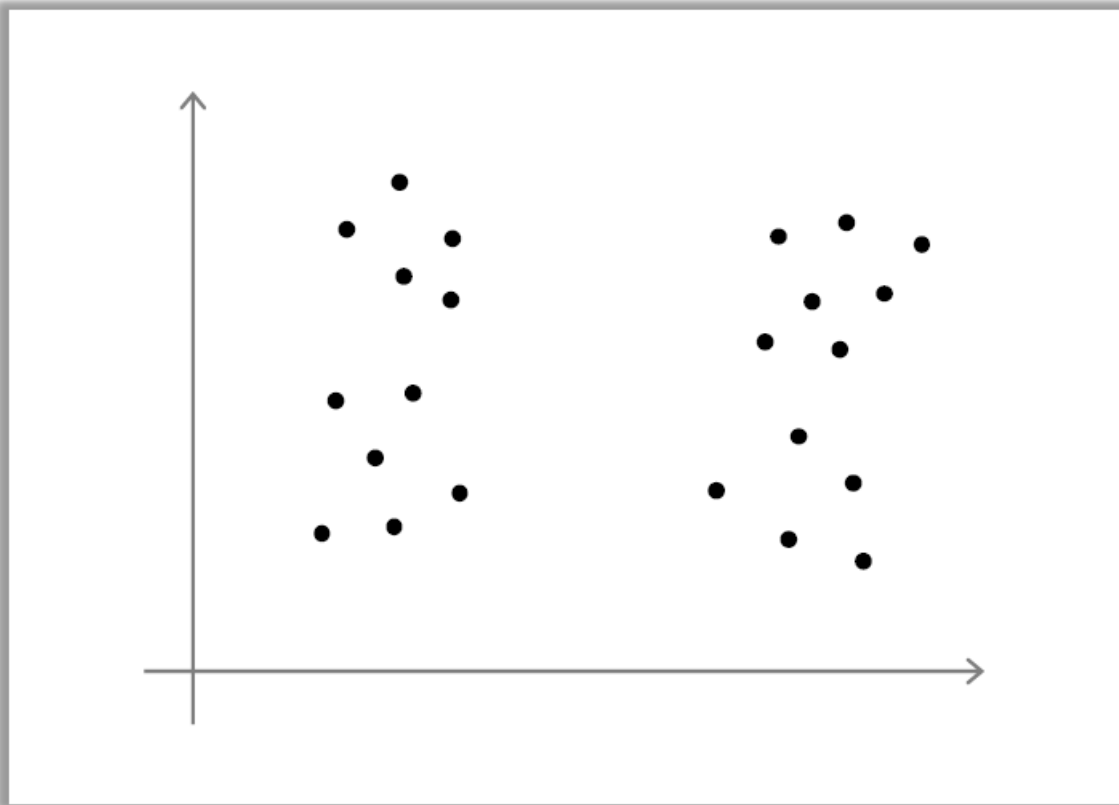
compute cost function  $J(c^{(1)}, c^{(2)}, \dots, c^{(m)}, \mu_1, \mu_2, \dots, \mu_k)$

}

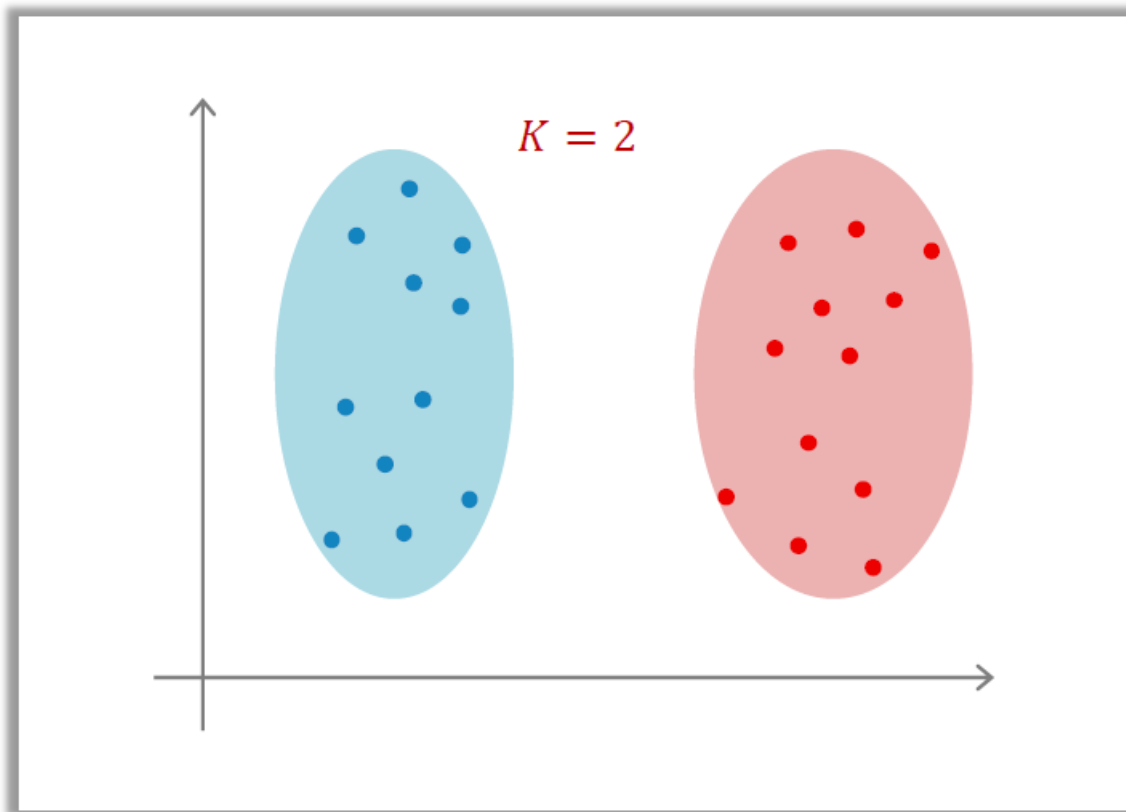
pick clustering with minimum cost

# تعیین تعداد خوشه‌ها

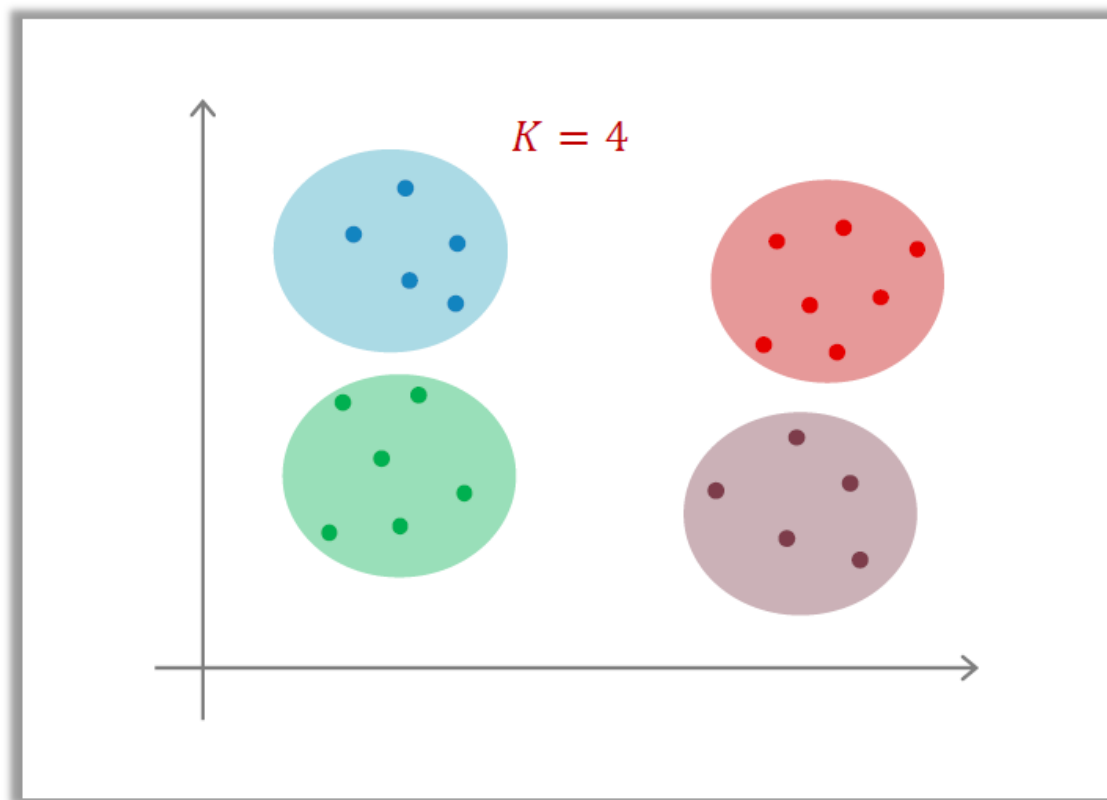
## فصل یازدهم : مقدار مناسب برای $K$ کدام است؟



## فصل یازدهم: مقدار مناسب برای $K$ کدام است؟

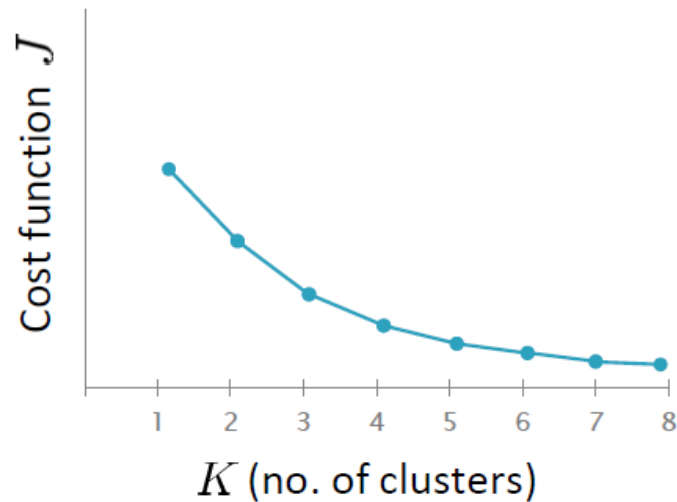
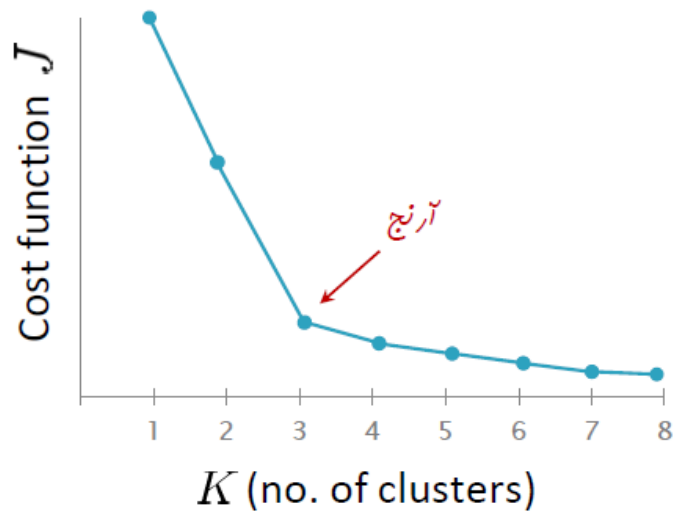


## فصل یازدهم: مقدار مناسب برای $K$ کدام است؟



## فصل یازدهم : تعیین تعداد مناسب خوشه ها

روش «آرنج» □



بهبود خوشه‌بندی

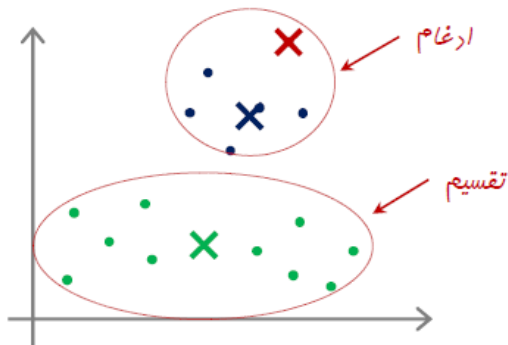


## فصل یازدهم : بهبود خوشه بندی با پس پردازش خوشه ها

□ تقسیم.

□ تقسیم یک خوشه با بیشترین خطا به دو خوشه

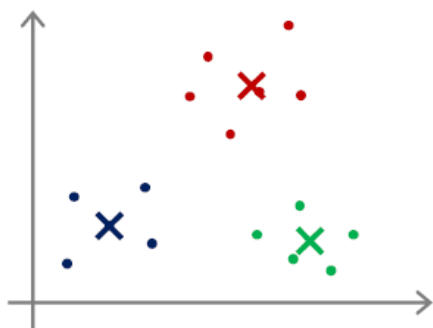
با اجرای K-means بر روی داده های  
این خوشه با مقدار  $K = 2$



□ ادغام.

□ ادغام نزدیک ترین دو خوشه

□ ادغام دو خوشه با حداقل افزایش در مجموع خطا



## فصل یازدهم: الگوریتم K-means دو بخشی ساز

□ الگوریتم دو بخشی ساز.

□ با یک خوشه شامل تمامی داده‌ها شروع کن.

□ هر بار یک خوشه را انتخاب کن:

■ خوشه‌ی انتخاب شده را به وسیله‌ی الگوریتم K-means به دو خوشه تقسیم کن.

■ مجموع خطای خوشه‌بندی را محاسبه کن.

■ خوشه‌بندی با کمترین خطا را انتخاب کن.

□ عمل فوق را تا زمان رسیدن به تعداد خوشه‌های مورد نظر تکرار کن.

## فصل یازدهم: الگوریتم K-means دو بخشی ساز

Start with all the points in one cluster

while the number of clusters is less than  $K$

    measure the total error

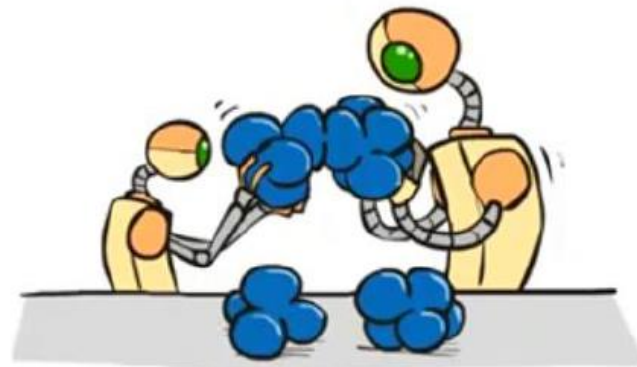
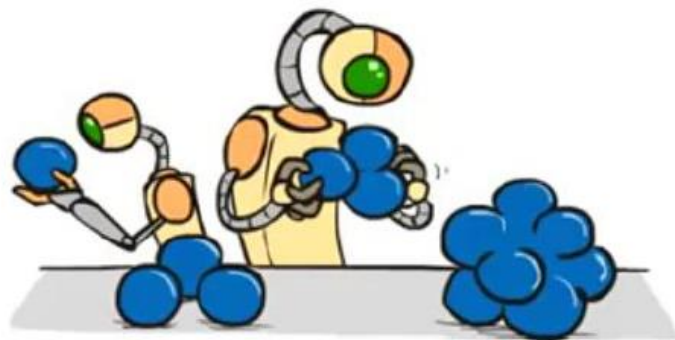
    for every cluster

        perform K-means clustering with  $k = 2$  on the given cluster

        measure the total error after splitting

    choose the cluster split that gives the lowest error

## فصل یازدهم : خوشه بندی سلسله مراتبی



## فصل یازدهم : خوشه بندی سلسله مراتبی

□ خوشه بندی سلسله مراتبی.

□ ابتدا داده های بسیار شبیه را ادغام کن.

□ به تدریج با ادغام خوشه های کوچک تر، خوشه های بزرگ تری ایجاد کن.

□ الگوریتم.

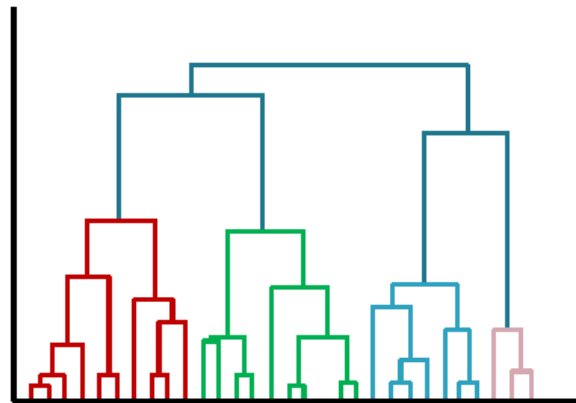
□ در ابتدا هر داده بیانگر یک خوشه است.

□ مراحل زیر را تکرار کن:

■ هر بار نزدیک ترین دو خوشه را انتخاب کن.

■ آن دو خوشه را در یک خوشه جدید ادغام کن.

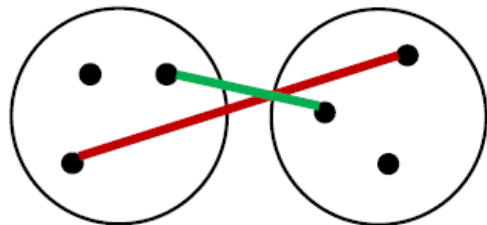
■ توقف: زمانی که تنها یک خوشه باقی مانده باشد.



□ ایجاد یک **درخت نگاره** شامل یک طیف گسترده از خوشه بندی ها.

## فصل یازدهم : خوشه بندی سلسله مراتبی

□ س. چگونه می توان نزدیک ترین دو خوشه را تعریف نمود؟



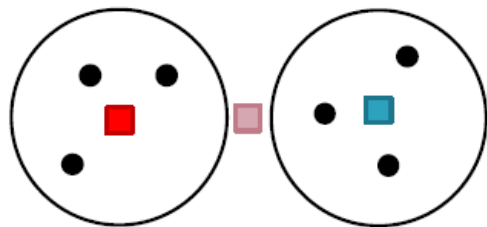
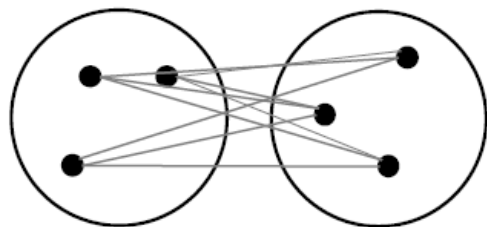
□ معیارهای تعیین شباهت خوشه ها.

□ **نزدیک ترین زوج** (خوشه بندی تک-پیوندی)

□ **دورترین زوج** (خوشه بندی تمام-پیوندی)

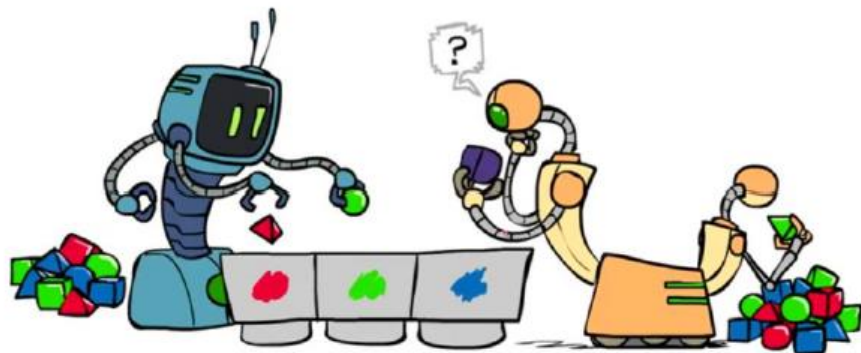
□ میانگین فاصله همه زوج ها

□ روش «وارد» (کمترین پراکندگی، مانند k-means)



□ معیارهای مختلف باعث ایجاد خوشه بندی های متفاوتی می شوند.

## فصل یازدهم : خلاصه



□ یادگیری بدون نظارت. یافتن ساختار در داده‌ها

□ خوشه‌بندی. گروه‌بندی داده‌های مشابه

□ الگوریتم خوشه‌بندی K-means

■ پیاده‌سازی آسان

■ برای مجموعه داده‌های بسیار بزرگ کند

■ امکان گیر کردن در بهینه‌ی محلی

□ پس‌پردازش خوشه‌ها: تقسیم و ادغام خوشه‌ها

□ الگوریتم K-means دو بخشی‌ساز

■ خوشه‌بندی بهتر نسبت به الگوریتم K-means

□ الگوریتم‌های خوشه‌بندی سلسله‌مراتبی

تمرین‌ها

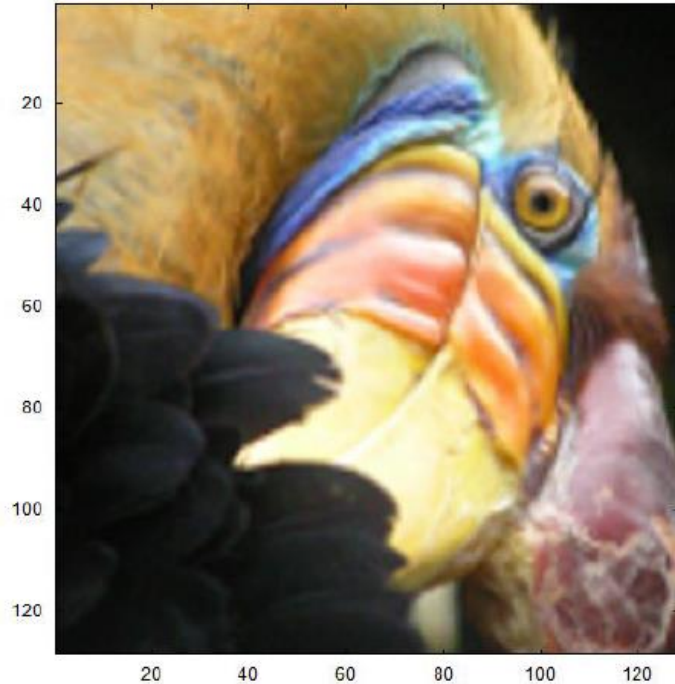


## فصل یازدهم: تمرین ۱ (پیاده سازی الگوریتم K-means)

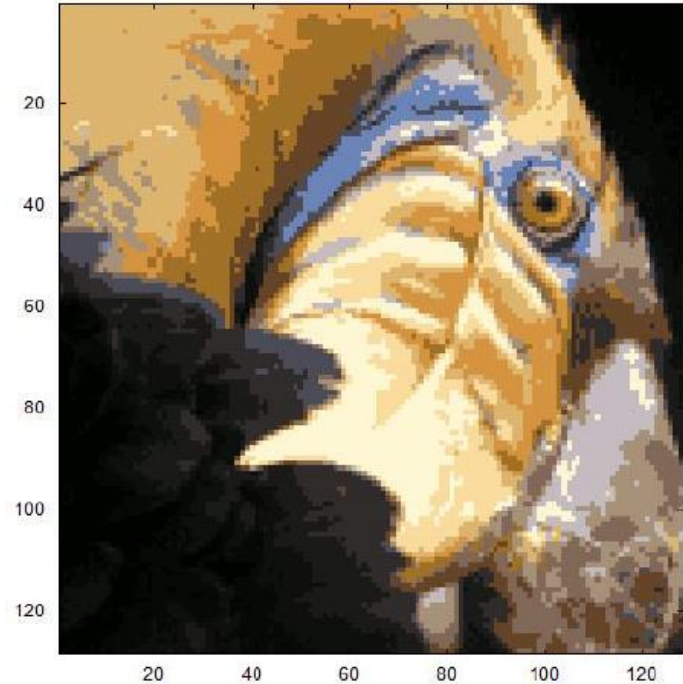


## فصل یازدهم : تمرین ۲ (فشرده سازی تصویر با استفاده از K-means)

تصویر اصلی



تصویر فشرده شده (۱۶ رنگ)



## فصل یازدهم :



باتشکر از توجه شما

