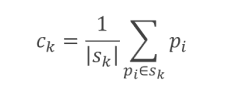
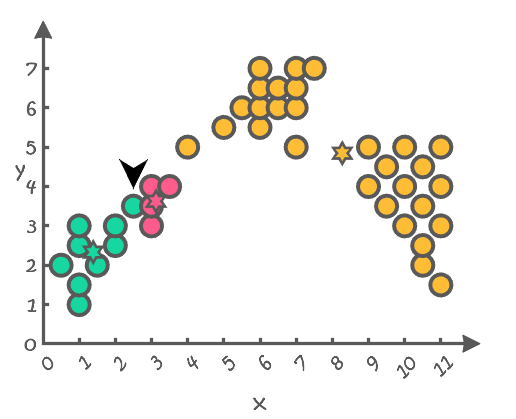
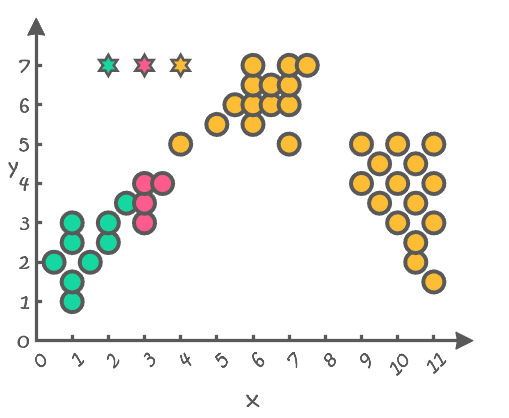
**الگوریتم K-Means**

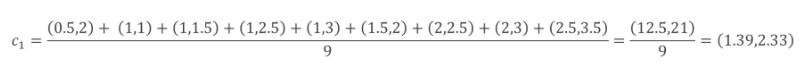
هدف : میخواهیم در نمودار دو بعدی یا به عبارت بهتر در اطلاعاتی با دو ویژگی را خوشه بندی کنیم

1. ابتدا تعیین مبکنیم که میخواهیم اطلاعات را به چند دسته تقسیم کنیم.
2. به تعداد خوشه هایمان نقاط رندوم در فضا در نظر میگیریم.
3. فاصله هر کدام از داده ها را نسبت به نقاط رندوم که انتخاب کردیم حساب می کنیم(برای این کار میتوانیم از فاصله منهتن استفاده کنیم و یا فاصله اقلیدسی).
4. فاصله داده با خوشه ها را مقایسه می کنیم ،هر کدام از داده های روی نمودار به نقاط رندوم یا به عبارتی به خوشه هایمان نزدیک تر بود جزو آن خوشه می شود و یا لیبل آن خوشه را می گیرد.
5. مرحله 4 برای تک تک داده‌ها انجام می‌شود.
6. نقاط رندوم را به نقاط میانگین تجمع داده ها می بریم.(با فرمول زیر)

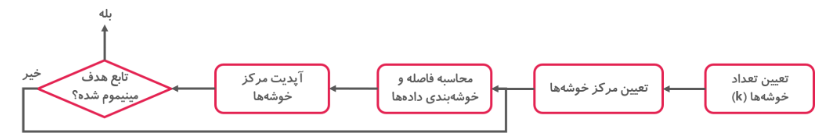


به عنوان مثال برای نمودار و نقاط سبز زیر داریم: (توپ = داده ، ستاره = خوشه)

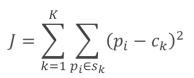
  



1. حال چون فاصله نقاط با خوشه ها عوض شده، باید مرحله 4 تا 6 را تا جایی تکرار کنیم که دیگر خوشه ها تغییر فاصله ندهند و نیز اعضا تغییر گروه ندهند.



تابع هدف در این الگوریتم:



توضیح: سیگمای داخلی فاصله داده‌های یک خوشه را با مرکز همان خوشه محاسبه می‌کند. زمانی J به کمترین مقدار می‌رسد که هر داده‌ای دقیقا در خوشه درست خودش قرار بگیرد. اگر داده 1 اشتباها در خوشه 3 باشد، نتیجه این می‌شود که مقدار J افزایش می‌یابد.