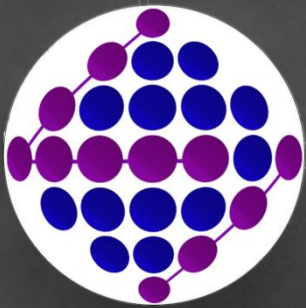


K Means Clustering



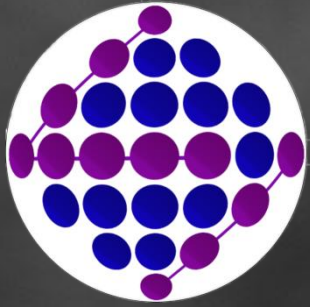
فصل دوازدهم



شرکت فروش مصنوعي رسا

مدرس : خريد فاشمي نژاد

K Means Clustering Theory



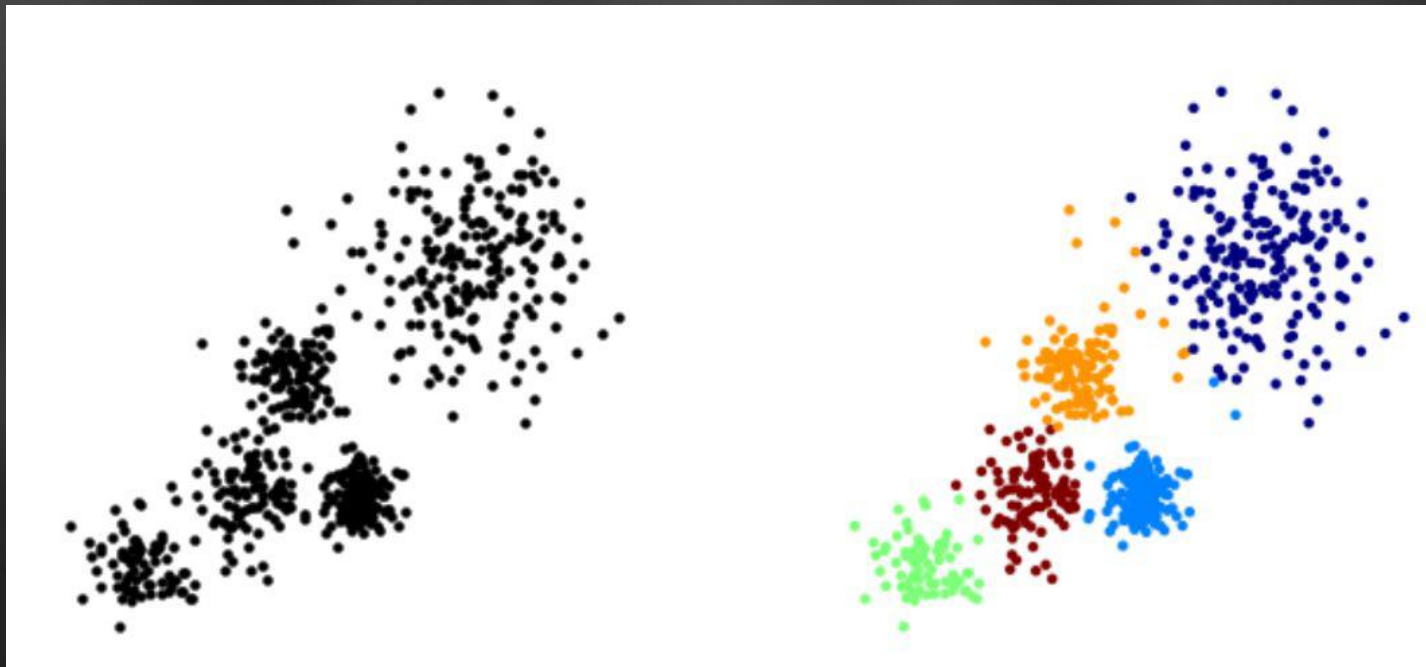
جلسہ سی و ہفتم



شرکت فروش مصنوعي رسا

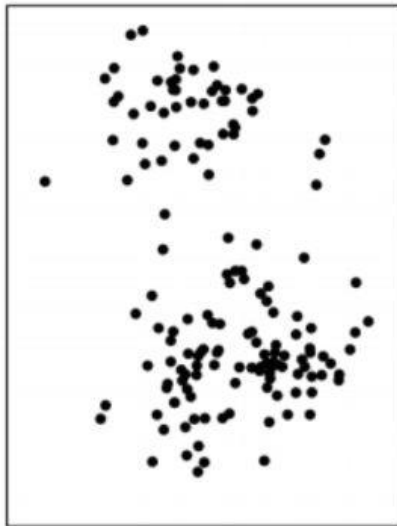
مدرس : فرید ہاشمی نژاد

• هدف کلی این است که داده ها را به گروه های متمایزی
تقسیم کنیم به گونه ای که مشاهدات درون هر گروه مشابه
باشد



- الگوریتم K به معنی انتخاب تعداد خوشه ها است. هر نقطه را به صورت تصادفی به یک خوشه اختصاص میدفیم تا زمانی که خوشه ها قابلیت تغییر داشته باشند، موارد زیر تکرار می شود:
- هر نقطه داده را به خوشه ای اختصاص دفید که مرکز آن نزدیکترین است.

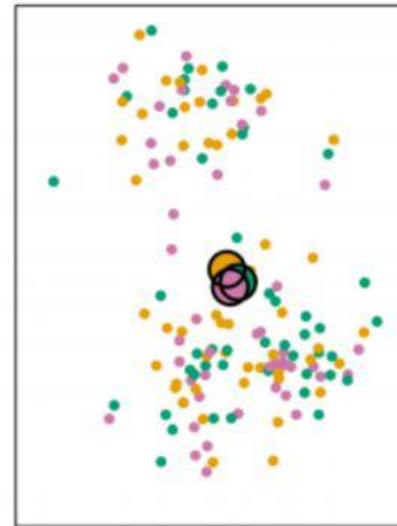
Data



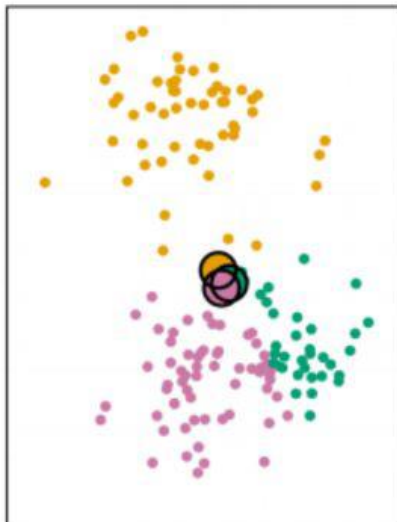
Step 1



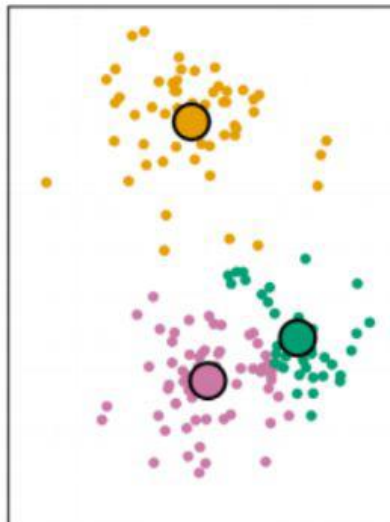
Iteration 1, Step 2a



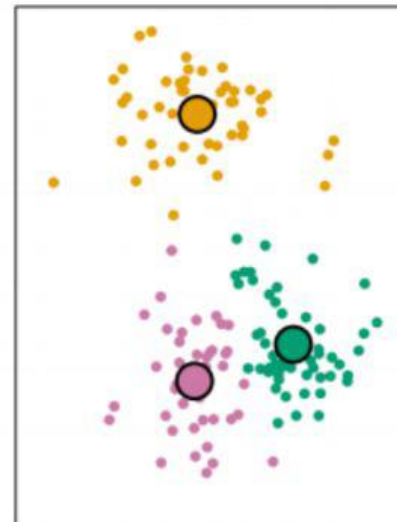
Iteration 1, Step 2b



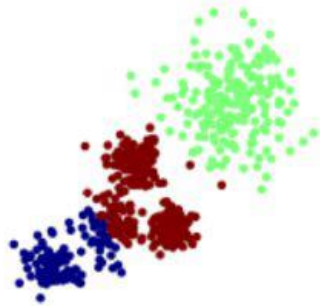
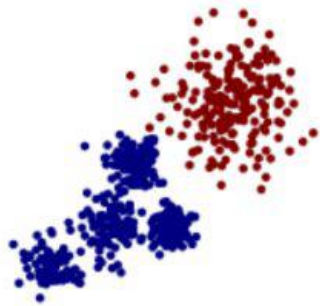
Iteration 2, Step 2a



Final Results



انتخاب K



برای انتخاب بهترین K هیچ پاسخ آسانی وجود ندارد
یک راه روش آرنج (elbow) هست اول از همه ، مجموع خطای مربع (SSE) را برای مقادیر مختلف K (برای مثال ۲ ، ۴ ، ۶ ، ۸ و...) محاسبه کنید .

SSE به عنوان مجموع مربع فاصله بین هر یک از اعضای خوشه و مرکز آن تعریف می شود. (sum of squared error)
اگر SSE را به ازای هر K ترسیم کنید ، خواهید دید که با بزرگتر شدن k خطا کاهش می یابد. و به این دلیل است که وقتی تعداد خوشه ها افزایش می یابد ، آنها باید کوچکتر شوند ، بنابراین اعوجاج نیز کوچکتر است. ایده روش آرنج این است که k را انتخاب کنید که در آن SSE به طور ناگهانی کاهش می یابد. این یک "اثر آرنج" در نمودار ایجاد می کند ، همانطور که در تصویر اسلاید بعد مشاهده می کنید

