

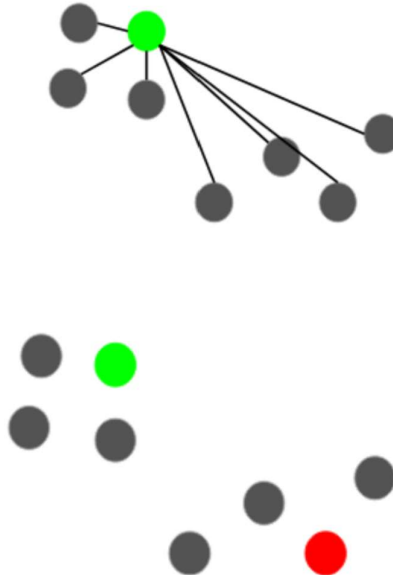
### سوال (۱)

الف) به نظرم چون ابعاد کم هست و داده‌ها هم برچسب دارن LDA خوب باشه.

ب) به نظرم چون داده‌ها category شده هستند پس احتمالا درخت تصمیم بهتر باشه (ولی مطمئن نیستم).

### سوال (۲)

توی k-means ما کلا نقطه‌های شروعمون رو تصادفی انتخاب می‌کردیم ولی توی k-means++ ما به تفاوتی توی شروع (initialization) انتخاب اون مرکزها داریم و کاری که می‌کنیم اینه که یکی از داده‌ها رو به عنوان مرکز انتخاب می‌کنیم و حالا برای انتخاب مرکز بعدی، یکی دیگه از داده‌ها که بیشترین فاصله را به مرکز قبلی داشته را انتخاب می‌کنیم که باعث میشه احتمالا این مراکز در خوشه‌های مختلفی قرار گرفته باشند و همینکه این مراکز در خوشه‌های مختلفی باشند (کیفیت بهتر) باعث می‌شود که سرعت همگرایی ما بیشتر شود.



### سوال (۳)

الف) غلطه و این گزاره درست نیست.

ب) غلطه چون همه چیزهایی که توی این الگوریتم داریم از اردر  $N^2$  هستند پس خودش هم همینه.

ج) درسته، این الگوریتم داده‌های پرت رو هم خودش تشخیص میده و اونا رو توی خوشه‌ی خاصی نمیذاره.

د) درسته، این الگوریتم خودش با کمک پارامترهاش، تعدادی خوشه تشکیل میده.