



بازیابی پیشرفته اطلاعات

نیم سال اول ۱۴۰۲-۰۳

مدرس: دکتر حمید بیگی

۴ دی

زمان: ۲۵ دقیقه

کوییز چهارم (۱۰۰ نمره)

سوال ۱: سوال شهودی (۳۰)

اگر از الگوریتم‌های distance-based مانند k-means برای خوشه‌بندی داده‌هایی که ابعاد بالایی دارند استفاده کنیم، چه مشکلی رخ می‌دهد؟

هنگام استفاده از الگوریتم ذکر شده، چگونه استفاده از *normalization* به روی داده‌ها می‌تواند موثر باشد؟ مثال بزنید.

مشکل الگوریتم k-means را در خوشه‌بندی داده‌ی عکس پایین تشریح کنید.



سوال ۲: Hierarchical (۳۰)

نقاط زیر داده شده است:

$$A = 1, B = 2, C = 3, D = 8, E = 9, F = 10$$

الگوریتم خوشه‌بندی سلسه مراتبی را به صورت پایین به بالا bottom-up انجام دهید و فرمول محاسبه فاصله را $d(x, y) = |x - y|$ در نظر بگیرید.

سوال ۳: K-means (۴۰)

نقاط زیر داده شده است:

$$A_1 = (2, 10), A_2 = (2, 5), A_3 = (8, 4), A_4 = (5, 8), A_5 = (7, 5), A_6 = (6, 4), A_7 = (1, 2), A_8 = (4, 9)$$

با فرض *centroid* قرار دادن نقاط اول، چهارم و هفتم، الگوریتم k-means را برای دو *epoch* انجام دهید.