دانشگاه صنعتی اصفهان
دانشکده مهندسی برق و کامپیوتر

عنوان: تکلیف سوم درس داده کاوی

نام و نام خانوادگی: نیلوفر سعیدی
شماره دانشجویی: ۹۸۲۲۹۶۳

1

f-measure $= 2 \times \frac{precision \times recall}{precision + recall}$
precision $= \frac{TP}{TP+FP}$
recall $= \frac{TP}{TP+FN}$
for last layer clusters: C1 = (p1,p2)
$P_{A1} = 1$, $P_{B1} = 0$
$R_{A1} = 0.67$, $R_{B1} = 0$
$F_{A1} = 0.8$, $F_{B1} = 0$
C2 =(p4,p5):
$P_{A2} = 0$, $P_{B2} = 1$
$R_{A2} = 0$, $R_{B2} = 0.4$
$F_{A2} = 0$, $F_{B2} = 0.57$
C3 =(p3,p6):
$P_{A3} = 0.5$, $P_{B3} = 0.5$
$R_{A3} = 0.33$, $R_{B3} = 0.2$
$F_{A3} = 0.4$, $F_{B3} = 0.29$
C4 = (p7,p8):
$P_{A4} = 0$, $P_{B4} = 1$
$R_{A4} = 0$, $R_{B4} = 0.4$
$F_{A4} = 0$, $F_{B4} = 0.57$

for second layer clusters:C1 = (p1,p2, p4, p5)
$P_{A1} = 0.5$, $P_{B1} = 0.5$
$R_{A1} = 0.67$, $R_{B1} = 0.4$
$F_{A1} = 0.57$, $F_{B1} = 0.44$
C2 = (p3,p6, p7, p8)
$P_{A2} = 0.25$, $P_{B2} = 0.75$
$R_{A2} = 0.33$, $R_{B2} = 0.6$
$F_{A2} = 0.28$, $F_{B2} = 0.67$

2

This occurs when no points are assigned to other centroids during the assignment step, the re-calculation step does not get rid of this cluster, and it also does not re-calculate the centroid value because no points are being used and so essentially we will have an output

with 1 cluster. The only solution is to choose a replacement centriod, this can be done by:

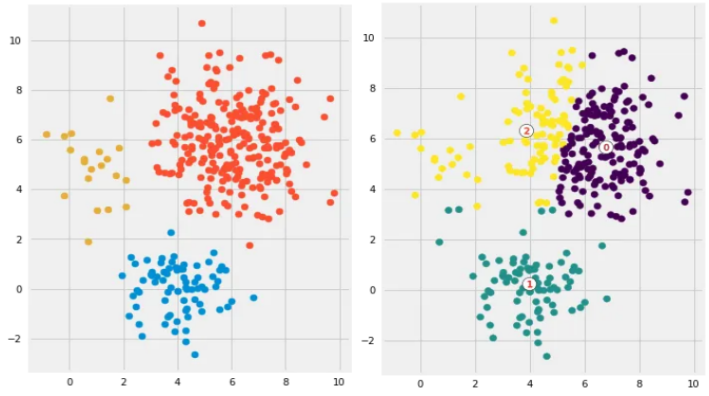
Repeat until no empty clusters:

Choosing the point that contributes most to SSE

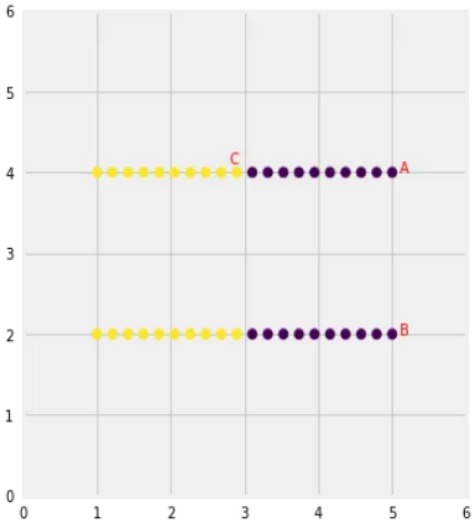
Choose a point from the cluster with the highest SSE

3

When multiple runs fixes it:



When multiple runs always results in the same wrong result:



4

4.1 comparison with k-medoids

Strength: K-means is easier and faster to compute.
Weakness: is much more susceptible to noise and outliers than k-medoids is.

4.2 comparison with hierarchical clustering

Strength: k-means is more efficient in terms of space and time complexity, and performs more accurately in specific shapes of clusters like multidimensional ovals or spheres.
Weakness: We need to know the amount of k in k-means, while in hierarchical clustering, we can choose between any number of clusters after the algorithm is done.