**yazan jarrar**

**problem a:**

the percentage change in cost after 10 iterations:

for c1 is:0.26483917144562047

for c2 is:0.7669795594605943



c2 is better than c1 because the distance between the points and their initial

centers will be big at initialization the points will be intense between the different center points when

the center point changes a lot more points will change which center they belong to so that will

give higher probability of finding better center point so the improvement rate will be better

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**problem b:**

the percentage change in cost after 10 iterations:

for c1 is: -0.02966459207764418

for c2 is: -0.0018895865097607058



c1 is better than c2 because in the cosine similarity the equation is divided over the length of each point

multiplied by the other so when the point is far the similarity will be higher so the points will get

in saturation faster so it will not improve a lot in c1 because the similarity will be lower at first the points

will be intense between the different center points when the center point changes

a lot more points will change which center they belong to so that will give higher probability of finding

better center point so the improvement rate will be better

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Euclidean vs Cosine for K-Mean:**

as we saw in problem a and problem b it depends on initial centers positions but overall, the cosine similarity will get to saturation faster and Euclidean will give better percentage change in cost.

