M. Taylor Long
Agata Miszczyk
CSCI 5715
HW 4

Homework 4

1. **Consider a dataset with following points in 2-dimensional space: (1,1), (2,1), (1,2), (2,2), (7,1), (8,1), (7,2), and (8,2). Show the clustering results using k-means with each of the following initial seed pairs.**

    a. seed 1: (3, 2)
    seed 2: (5, 2)

    First Iteration:
    seed 1: (1,2),(2,2),(1,1),(2,1) → Mean Center (1.5,1.5)
    seed 2: (7,2),(8,2),(7,1),(8,1) → Mean Center (7.5,1.5)

    Second Iteration:
    No Change

    b. seed 1: (3, 1)
    seed 2: (3, 0)

    First Iteration:
    seed 1: (1,2),(2,2),(1,1),(2,1),(7,2),(8,2),(7,1),(8,1) → Mean Center (4.5, 1.5)
    seed 2: nothing! → Mean Center (3,0)

    Second Iteration:
    seed 1: (1,2),(2,2),(1,1),(2,1) → Mean Center (1.5,1.5)
    seed 2: (7,2),(8,2),(7,1),(8,1) → Mean Center (7.5,1.5)

    Third Iteration:
    No Change

    c. seed 1: (4, 1.5)
    seed 2: (15, 1.5)

    First Iteration:
    seed 1: (1,2),(2,2),(1,1),(2,1),(7,2),(8,2),(7,1),(8,1) → Mean Center (4.5, 1.5)
    seed 2: nothing! → Mean Center (15, 1.5)

    Second Iteration:
    seed 1: (1,2),(2,2),(1,1),(2,1),(7,2),(8,2),(7,1),(8,1) → Mean Center (4.5, 1.5)
    seed 2: Nothing! → Mean Center (15, 1.5)
    No change from first

    Third Iteration:
    Keeps looking for second cluster, but there isn't one.
    We assume it keeps repeating and looping through the set of points but never finds a second cluster and therefore goes on forever looking for K = 2 or eventually gives up.

M. Taylor Long
Agata Miszczyk
CSCI 5715
HW 4

2. Compute participation index (an interest measure for colocation) for the following pairs:

a. (Burger King, Pizza Hut)
pi(BK,PH) = min { pr(BK, (BK,PH)) , pr(PH, (BK,PH)) }
pr(BK, (BK,PH)) = 1/10
pr(PH, (BK,PH)) = 1/4
Therefore...
pi(BK,PH) = 1/4

b. (Pizza Hut, McDonalds)
pr(PH,(PH,MC)) = 3/4
pr(MC,(PH,MC)) = 3/13
pi(PH,MC) = 3/13

c. (Burger King, McDonalds)
pr(BK,(BK,MC)) = 6/10
pr(MC,(BK,MC)) = 6/13
pi(BK,MC) = 6/13