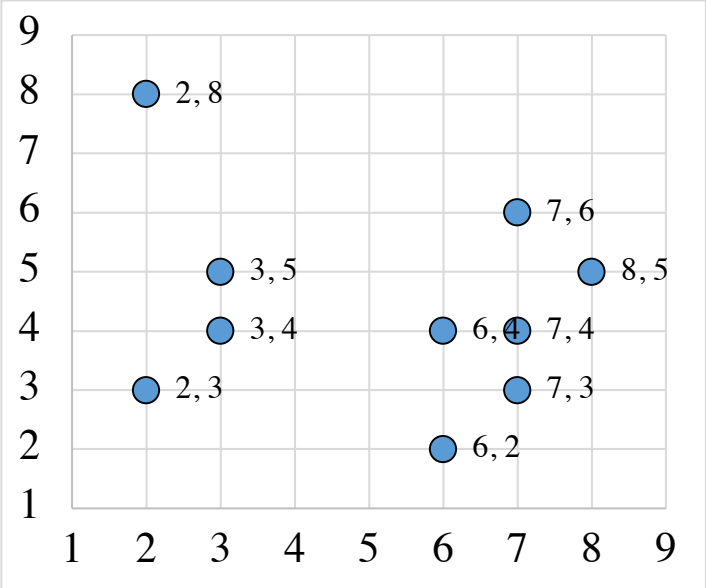
CSE 40647/60647 Data Science (Spring 2018)
Lecture 17: Clustering: K-Partitioning Methods

Suppose we have 10 college football teams X1 to X10. We want to cluster them into 2 groups. For each football team, we have two features: One is # wins in Season 2016, and the other is # wins in Season 2017.

Team	# wins in Season 2016 (x-axis)	# wins in Season 2017 (y-axis)
X1	3	5
X2	3	4
X3	2	8
X4	2	3
X5	6	2
X6	6	4
X7	7	3
X8	7	4
X9	8	5
X10	7	6



(1) Initialize with two centroids, (4, 6) and (5, 4). Use Manhattan distance as the distance metric. Please use K-Means to find two clusters.

(2) Initialize with two centroids, (4, 6) and (5, 4). Use Euclidean distance as the distance metric. Please use K-Means to find two clusters.

(3) Initialize with two centroids, (3, 3) and (8, 3). Use Manhattan distance as the distance metric. Please use K-Means to find two clusters.

(4) Initialize with two centroids, (3, 2) and (4, 8). Use Manhattan distance as the distance metric. Please use K-Means to find two clusters.

(5) Suppose we initialize with two medoids, (2, 8) and (6, 2). Use Euclidean distance as the distance metric. In K-Medoids clustering, given a non-medoid (3, 5), do we swap the medoid (2, 8) with (3, 5)?

What are your observations?

Name (NetID):

What is the advantage of K-Means?

What are the disadvantages of K-Means?

What is the disadvantage of K-Medoids?