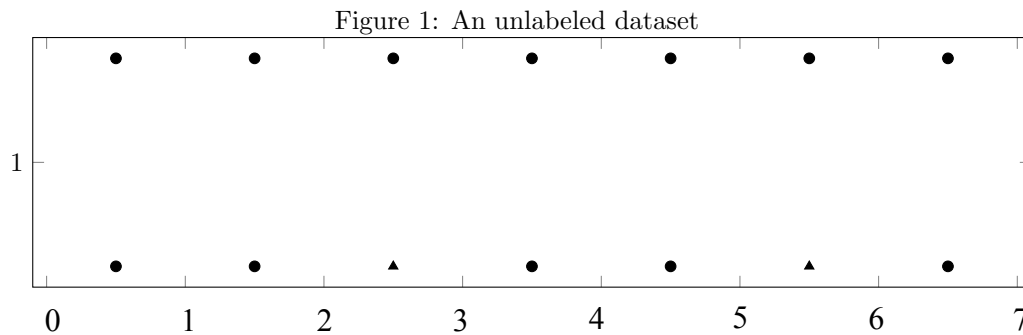


## Homework4 — Clustering

Consider the following dataset. All points are unlabeled and part of the same set. The triangles are used to distinguish two points later. *Please do not draw on this diagram until you have read the problems below.*

Suppose we run the K-means algorithm on this dataset with  $K = 2$  and the two points indicated by triangles as the initial centroids. When the algorithm converges, there will be two clearly separated clusters. Directly on Figure 1, draw a straight line that separates these two clusters, as well as the centroids of these two clusters.



Next, find two other different sets of initialize centroids that will converge to the exact same result if we apply K-means. Please follow the instructions below

- the initialize centroids have to be points of the dataset;
- directly on Figure 2, use two triangles to indicate the first set, and two squares to indicate the second;
- these two sets of points can overlap with each other, but of course cannot be the same;
- similarly these sets can overlap with the initialization of Figure 1 but cannot be the same;
- do not pick those that lead to ambiguous results due to different ways of breaking ties.

