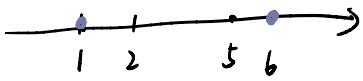
Name: Jieying Gong **Student ID:** 1009388066

CSCC11 Worksheet L11B KMeans

Q1 Consider a 1-dimensional dataset with the following points: 1, 2, 5, 6. Using K-means clustering with K=2 and initial centroids at 1 and 6, what will be the smaller centroid after one iteration (one assignment and one update step)?

Answer: 1.5 & 5.5



Q2 In the KMeans++ initialization method for KMeans clustering, how are the initial centroids selected¿

- ☐ (a) All centroids are randomly selected from the dataset.
- ☐ (b) The first centroid is selected randomly, and each subsequent centroid is selected with a probability proportional to the Euclidean distance from the nearest already selected centroid.
- ☐ (c) Use the density of data points in different regions to select centroids.
- ☐ (d) Select centroids as close to each other as possible to minimize intra-cluster variance.
- ☑ (e) None of the above