Exercise 1 - k-Means Clustering

Figure 1 shows a data set consisting of 6 instances (circles), in the two dimensional space defined by two numerical attributes.

i Apply k-means clustering to this set, using the star-shaped points as initial cluster representative instances. Five additional pictures are given, all with the data points included. Use these to graphically show the process of determining the two clusters.

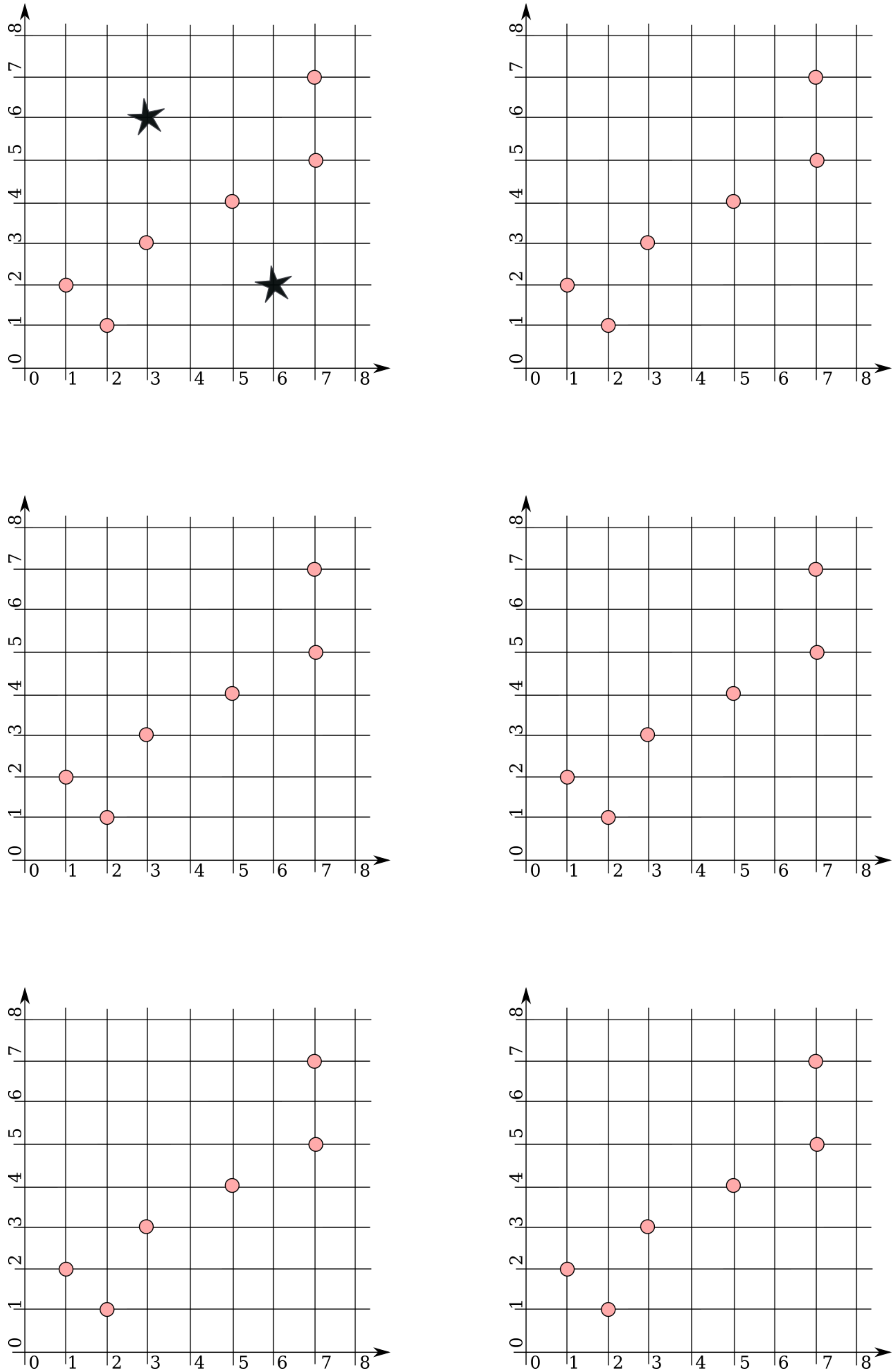
ii Clearly state the final cluster centroids and data instance membership in the clusters.

iii The within-cluster variance for some values of k and the same data set is as follows:

k	WCV
1	9.36
3	0.92
4	0.58

Calculate the within-cluster variance for $k = 2$ and determine what the value of k should be if the criterion is $WCV(k)/WCV(k-1) < 0.5 < WCV(k+1)/WCV(k)$.

Figure 1



Exercise 2 - k-Means Clustering

The upper left diagram of Figure 2 shows values of attributes x and y for a data set of 6 instances (A, B, D, E, F and G). If points C1 and C2 are initial randomly chosen centroids, describe and apply the k-means algorithm for clustering, explaining each step and what determines the end of the process. Draw each iteration in a separate diagram, using the grids provided. What is the final assignment of instances to clusters? What are the final cluster centroids?

Figure 2