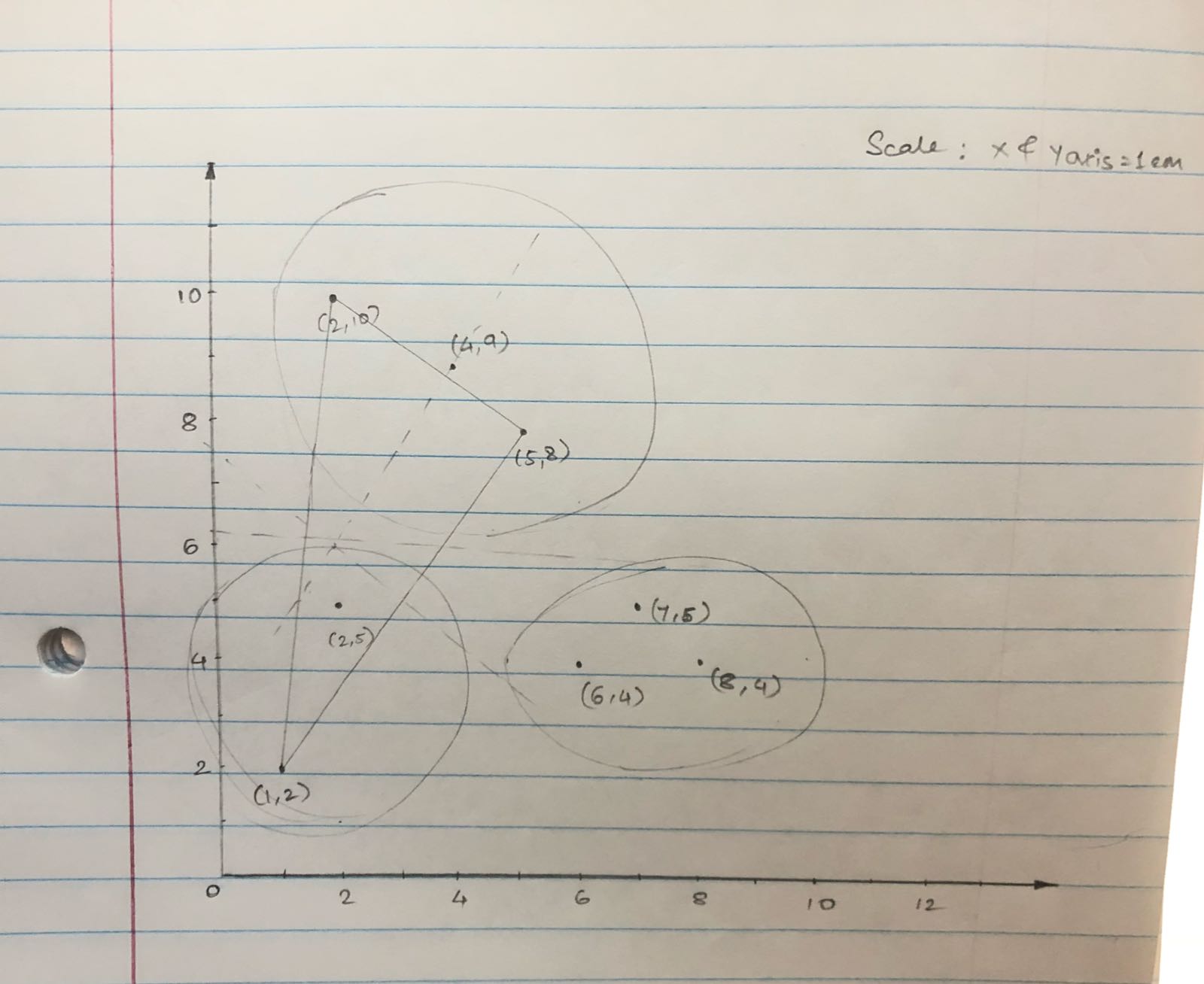
**Homework 7**

**Question 1:** Consider the following eight points in 2-dimensional space: (2,10); (2,5); (8,4); (5,8); (7,5); (6,4); (1,2); (4,9). Suppose we plan to use the Euclidean distance metric and that we are interested in clustering these points into 3 clusters.

1. Plot the data by hand on paper to see what might be appropriate clusters.
2. Beginning with the points (2,10), (5,8) and (1,2) as initial cluster centers, form the three initial clusters.
3. Use the 3-means clustering algorithm to get the final three clusters. What are the resulting centers and resulting clusters?
4. Mark the initial and final clusters on your graph and comment on what you see.

**Answer:**



**Cluster 1 consists of (1,2) & (2,5)**

**Cluster 2 consists of (2,10) , (5,8) & (4,9)**

**Cluster 3 consists of (6,4) , (7,5) & (8,4)**

**Question 2:**  k-means and EM clustering are available in the Weka Explorer. Both of them produce measures of the goodness of the clustering. For the simplified Iris and Glass datasets, graph the values of these measures for k-means and EM clustering from k = 1 to k = 5 clusters for both datasets (4 graphs total). What value of k seems best in each case?

**Answer:**

**DataSet: Iris**

**Clustering: K-mean**

**Value of K tested: 1, 2, 3, 4, 5**

**Best Case: k = 3**

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**DataSet: Iris**

**Clustering: EM**

**Values tested for K: 1, 2, 3, 4, 5**

**Best Case: k = 3**

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a computer screen

Description automatically generated**

**DataSet: Glass**

**Clustering: K-mean**

**Value of K tested: 1, 2, 3, 4, 5**

**Best Case: k = 4**

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**DataSet: Glass**

**Clustering: EM**

**Values of K tested: 1, 2, 3, 4, 5**

**Best Case: 4**

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a computer

Description automatically generated**