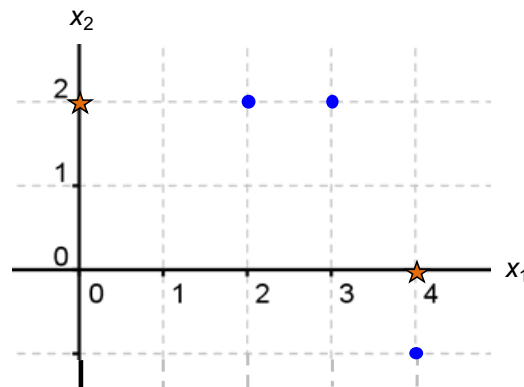


**K-means clustering**

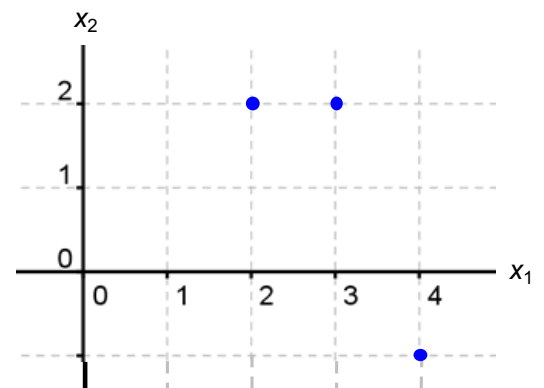
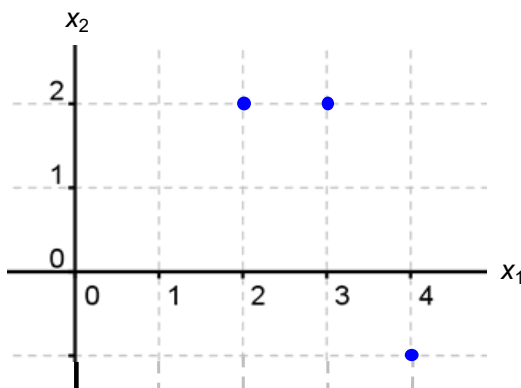
*(find and work with a partner)*

1. Consider the data below with  $n = 3$  and  $p = 2$ . The graph below shows these 3 points (circles), as well as the initial means (stars) for  $K = 2$ . Here  $\vec{\mu}_1^{(1)} = [0, 2]$  and  $\vec{\mu}_2^{(1)} = [4, 0]$ .

$$\mathbf{X} = \begin{bmatrix} 3 & 2 \\ 2 & 2 \\ 4 & -1 \end{bmatrix}$$



- (a) On the graph above, show the cluster membership of each point, based on these initial means. What are  $\mathcal{C}_1^{(1)}$  and  $\mathcal{C}_2^{(1)}$ ?
- (b) Based on these cluster memberships, what are  $\vec{\mu}_1^{(2)}$  and  $\vec{\mu}_2^{(2)}$ ? Draw these two points as stars on the left plot below. This concludes the first iteration of the  $K$ -means algorithm.



- (c) Based on the new means, draw the new cluster memberships and list  $\mathcal{C}_1^{(2)}$  and  $\mathcal{C}_2^{(2)}$ . Finally, on the right plot above, draw the final means  $\vec{\mu}_1^{(3)}$  and  $\vec{\mu}_2^{(3)}$  and write out their values.

