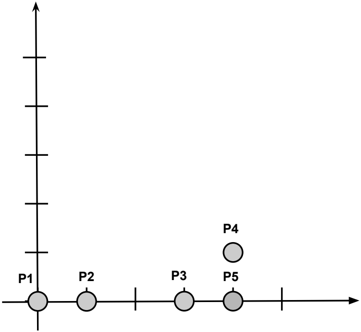
Exercise: K-Means Algorithm

Mauro Sozio

March 27, 2017

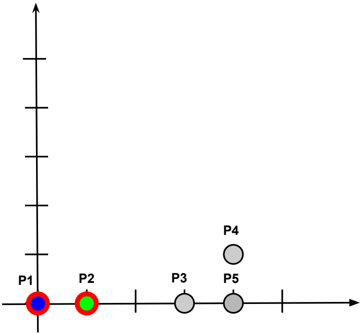
Input Points

Input: $P1 = (0,0), P2 = (1,0), P3 = (3,0), P4 = (4,1), P5 = (4,0)$.



Input Points and Centroids

We assume $P1$ and $P2$ are selected as centroids (denoted as $C1, C2$, respectively).



Reclustering-step 1

For each point $P3, P4, P5$, we determine the closest centroid

$$d(P3, P1) = 3, d(P3, P2) = 2, \text{P3 is assigned to the geen cluster}$$
$$d(P4, P1) = \sqrt{4^2 + 1^2} = \sqrt{17}, d(P4, P2) = \sqrt{10}, \text{P4 to the green cluster}$$
$$d(P5, P1) = 4, d(P5, P2) = 3, \text{P5 is assigned to the green cluster}$$

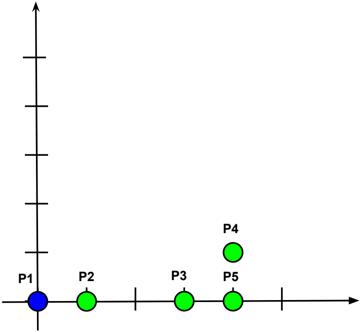
Reclustering-step 1

For each point $P3, P4, P5$, we determine the closest centroid

$$d(P3, P1) = 3, d(P3, P2) = 2, \text{P3 is assigned to the geen cluster}$$

$$d(P4, P1) = \sqrt{4^2 + 1^2} = \sqrt{17}, d(P4, P2) = \sqrt{10}, \text{P4 to the green cluster}$$

$$d(P5, P1) = 4, d(P5, P2) = 3, \text{P5 is assigned to the green cluster}$$

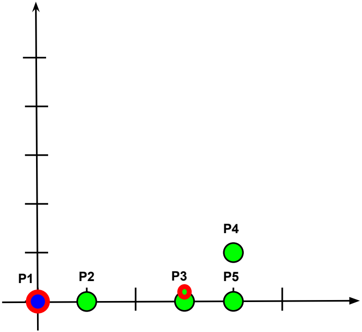


Recomputing centroids

We recompute the centroids $\bar{C}1$, $\bar{C}2$ as follows:

$$\bar{C}1 = P1$$

$$\bar{C}2 = \frac{P2 + P3 + P4 + P5}{4} = \left(\frac{1+3+4+4}{4}, \frac{1}{4}\right) = \left(3, \frac{1}{4}\right)$$



Reclustering-Step 2

For each point, we determine the closest centroid:

$$d(P2, \bar{C}1) = 1, d(P2, \bar{C}2) = \sqrt{(3-1)^2 + \frac{1}{16}} > 2, \text{P2 gets blue}$$

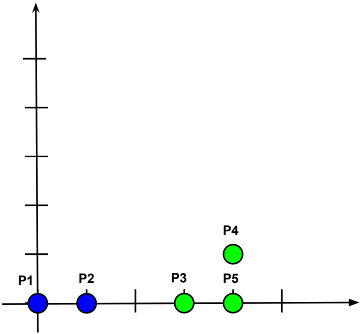
$$d(P3, \bar{C}1) = 3, d(P3, \bar{C}2) = \frac{1}{4}, \text{P3 remains green}$$

$$d(P4, \bar{C}1) = \sqrt{17}, d(P4, \bar{C}2) = \sqrt{1 + \frac{9}{16}} < 2, \text{P4 remains green}$$

$$d(P5, \bar{C}1) = 4, d(P5, \bar{C}2) = \sqrt{1 + \frac{1}{16}} < 2, \text{P5 remains green}$$

Reclustering-step 2

We obtain the following clustering:

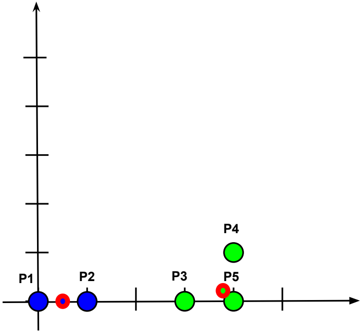


Recomputing the centroids

We recompute the centroids $\hat{C}1$, $\hat{C}2$ as follows:

$$\hat{C}1 = \frac{P1 + P2}{2} = \left(\frac{1}{2}, 0\right)$$

$$\hat{C}2 = \frac{P3 + P4 + P5}{3} = \left(\frac{3+4+4}{3}, \frac{1}{3}\right) = \left(\frac{11}{3}, \frac{1}{3}\right)$$



Reclustering-step 3

$$d(P1, \hat{C}1) = \frac{1}{2}, d(P1, \hat{C}2) > \frac{11}{3} > 3, \text{P1 stays blue}$$

$$d(P2, \hat{C}1) = \frac{1}{2}, d(P2, \hat{C}2) = \sqrt{\frac{4}{9} + \frac{1}{9}} > \frac{2}{3} > \frac{1}{2}, \text{P2 stays blue}$$

$$d(P3, \hat{C}1) > 2, d(P3, \hat{C}2) = \sqrt{\left(3 - \frac{11}{3}\right)^2 + \frac{1}{9}} < 1, \text{P3 stays green}$$

$$d(P4, \hat{C}1) > 2, d(P4, \hat{C}2) = \sqrt{\left(\frac{12}{3} - \frac{11}{3}\right)^2 + \frac{4}{9}} = \sqrt{\frac{5}{9}} < 1, \text{P4 green}$$

$$d(P5, \hat{C}1) > 2, d(P5, \hat{C}2) = \sqrt{\frac{2}{9}} < 1, \text{P5 stays green}$$

Final Clustering

As the clustering did not change, the algorithm terminates with the following final clustering: