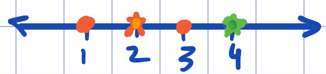
Micol Altomare

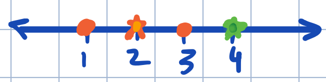
Problem 2

Initialization and first assignment (colour-coded) of the dataset.



$\mu_1 = 2 \; : x^{(1)}, x^{(2)}, x^{(3)}$
$\mu_2 = 4 \; : x^{(4)}$

During the moving-the-centroids step, the mean of the points within their respective cluster does not change, resulting in the points being assigned to their same centroids.



$\mu_1 = 2 : x^{(1)}, x^{(2)}, x^{(3)}$
$\mu_2 = 4 : x^{(4)}$

Consequently, the distortion is:

$$
\begin{aligned}
J &= \frac{1}{m} \sum_{i=1}^{m} \| x^{(i)} - \mu_{c^{(i)}} \|^2 \\
&= \frac{1}{4} \left[(1-2)^2 + (2-2)^2 + (3-2)^2 + (4-4)^2 \right] \\
&= \frac{1}{2}
\end{aligned}
$$

Meanwhile, a different initialization (and resulting cluster assignments) would be:

$\mu_1 = {}^1/_2 \; : x^{(1)}, x^{(2)}$
$\mu_2 = {}^7/_2 \; : x^{(3)}, x^{(4)}$

With distortion:

$$
\begin{aligned}
J' &= \frac{1}{m} \sum_{i=1}^{m} \| x^{(i)} - \mu_{c^{(i)}} \|^2 \\
&= \frac{1}{4} \Big[\left(1 - \tfrac{3}{2}\right)^2 + \left(2 - \tfrac{3}{2}\right)^2 + \left(3 - \tfrac{7}{2}\right)^2 + \left(4 - \tfrac{7}{2}\right)^2 \\
&= \frac{1}{4}
\end{aligned}
$$

$J' < J$ thus it's possible for k-means to converge to a solution that is not globally optimal.