Property Demonstration on the Clustering Algorithm with Various Examples

陈子蔚 Chen Ziwei

Department of Computer Science and Engineering

Recent Research Topic: Generative Model

Supervisor: 郑锋

Task 7-3

Consider all points

$$\mathcal{S} = \{s_1, \cdots, s_n\} \subset \mathbb{R}^2$$
$$|\mathcal{S}| = n = 128$$
$$s_1, \cdots s_{\frac{n}{2}} \sim \mathcal{N}([0.5,0.5]^\top, 0.5I)$$
$$s_{\frac{n}{2}+1}, \cdots, s_n \sim \mathcal{N}([-0.5,-0.5]^\top, 0.5I)$$

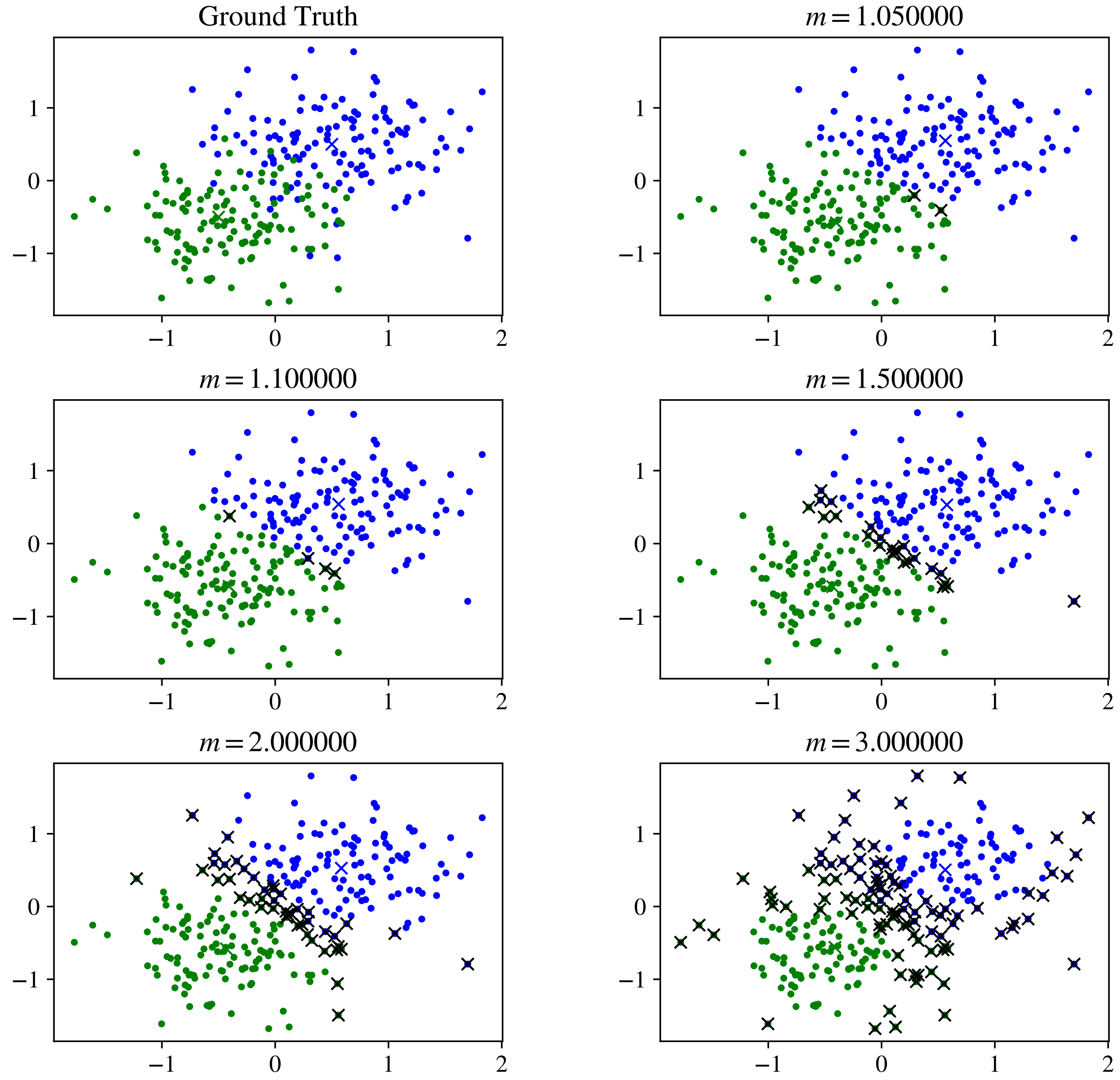
, which requiring clustering into $k = 2$ parts.

Various m, [1.05,1.1,1.5,2,3] were experimented in fuzzy c-means algorithm and the max number of iterations is 25.

The initial centers are the same in experiments.

Obviously, when the m increases, the number of the points with a black cross increase. A black cross on the point s_i indicates that $\max_j \mu_{ij}$, where ϵ is set to 0.7 here.

In other words, as m increases, the boundaries between the clusters become more soft, where points nearby boundaries would be easy to switch another cluster.



Task 7-4

Consider all points

$$\mathcal{S} = \{s_1, \cdots, s_n\} \subset \mathbb{R}^2$$

$$|\mathcal{S}| = n = 128$$

$$s_1, \cdots s_{\frac{n}{2}} \sim \mathcal{N}\left([0.5, 0.5]^\top, \mathrm{diag}([0.3, 0.5]^\top)\right)$$

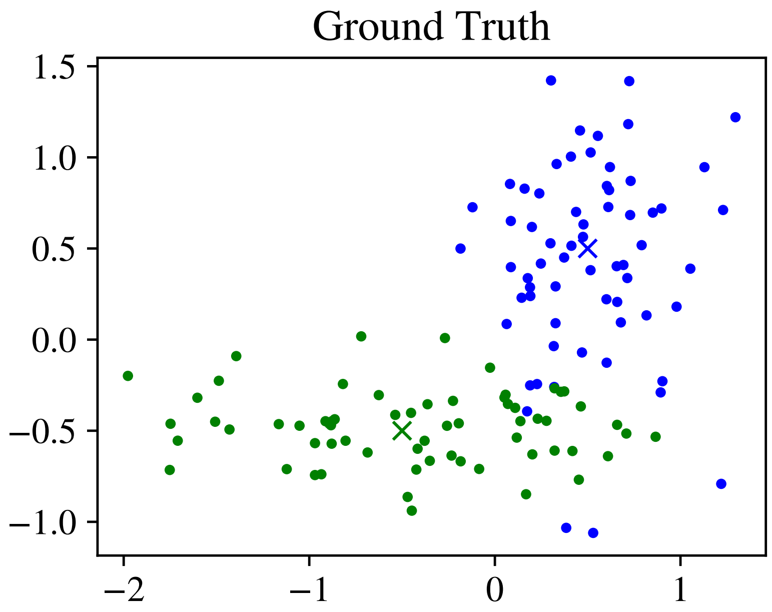
$$s_{\frac{n}{2}+1}, \cdots, s_n \sim \mathcal{N}\left([-0.5, -0.5]^\top, \mathrm{diag}([0.3, 0.5]^\top)\right)$$

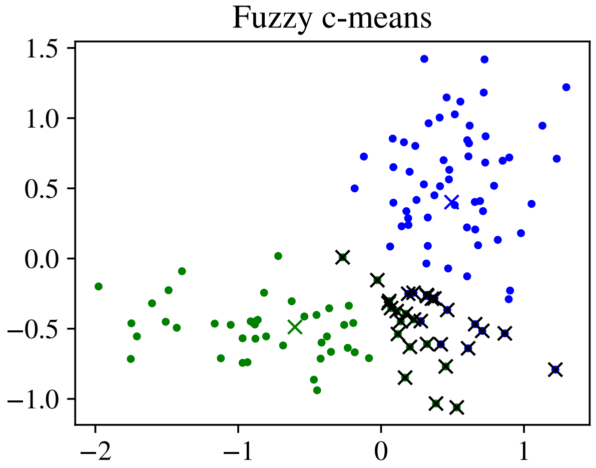
, which requiring clustering into $k = 2$ parts.

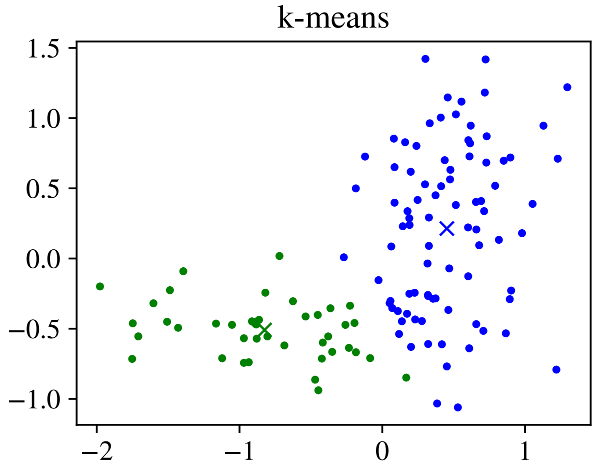
The initial centers are the same in experiments.

In the figure of the result from fuzzy c-means, a black cross on the point s_i indicates that $\max_j \mu_{ij}$, where ϵ is set to 0.7 here, and m is set to 2.

Obviously, the boundaries between the clusters in fuzzy c-means algorithm are softer, while the boundaries between the clusters in k-means algorithm are much harder with the comparison with the ground truth.





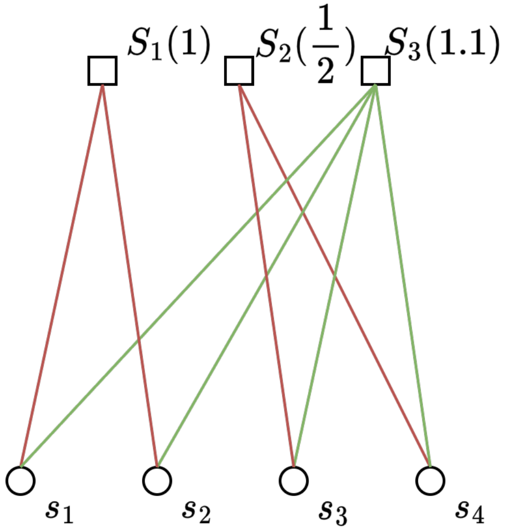


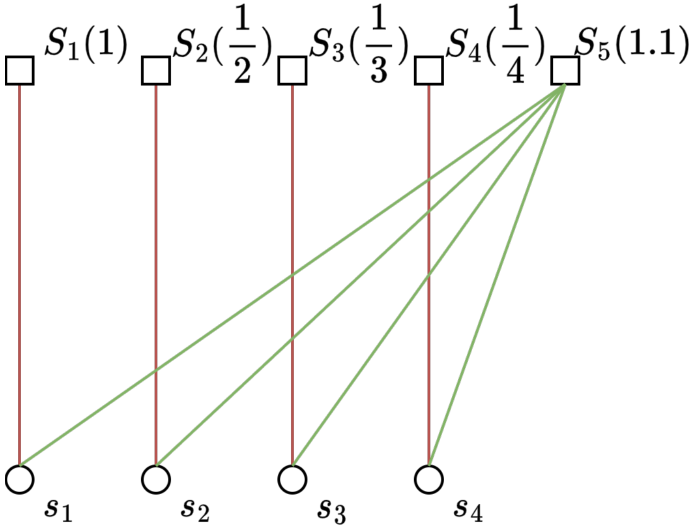
Task 8-1

These two figures show two examples, where subsets in the red indicate the solution obtained by greedy algorithm and subsets in the green indicate the optimal solution.

For the upper case, $\frac{w}{w^*} = \frac{1+\frac{1}{2}}{1.1} = 1.364$

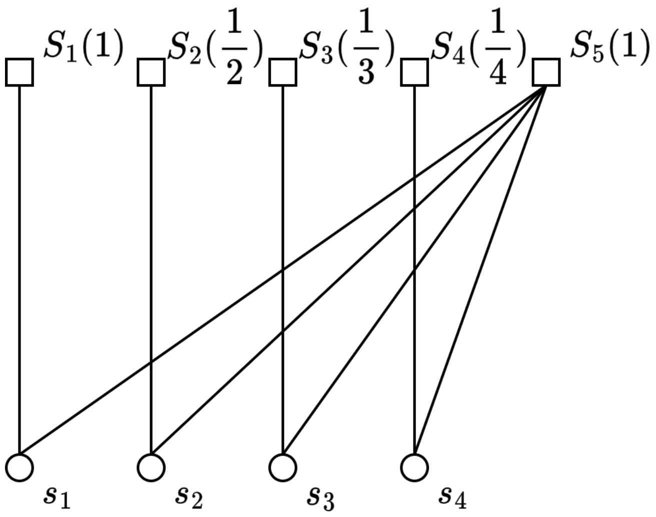
For the lower case, $\frac{w}{w^*} = \frac{1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}}{1.1} = 1.894$





Task 8-2

For the upper case, in the first step, S_4 and S_5 have the same average weight $\frac{1}{4}$ among the uncovered elements. If S_5 is chosen, $w^* = 1$ would be obtained. When S_1 is chosen, similar situations would be encountered.
Then, when S_1, S_2, S_3, S_4 are chosen, $w = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = H(4)w^*$ would be obtained, and when S_4, S_5 are chosen, $w = 1 + \frac{1}{4} = \frac{5}{4}$, where $w^* < w < H(4)w^*$.



For the upper case, in the first step, S_2, S_4 and S_5 have the same average weight $\frac{1}{4}$ among the uncovered elements. If S_5 is chosen, $w^* = 1$ would be obtained. When S_1 is chosen, similar situations would be encountered.
Then, when S_1, S_2, S_3, S_4 are chosen, $w = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = H(4)w^*$ would be obtained, and when S_1, S_2, S_3 are chosen, $w = 1 + \frac{1}{2} + \frac{1}{3}$, where $w^* < w < H(4) = H(4)w^*$.