

1. Which of the following statements best describes the iterative part of the K-means algorithm?

1 / 1 point

- ☐ The k-means algorithm assigns a number of clusters at random.
- ☒ The k-means algorithm adjusts the centroids to the new mean of each cluster, and then it keeps repeating this process until no example is assigned to another cluster.

Correct! Once the cluster centroids don't move anymore, it means that we have reached convergence.

- ☐ The k-means algorithm iteratively deletes outliers.
- ☐ The k-means algorithm iteratively calculates the distance from each point to the centroid of each cluster.

2. Which statement describes better “the smarter initialization of K-mean clusters?”

1 / 1 point

- ☐ “Draw a line between the data points to create 2 big clusters.”
- ☐ “After we find our centroids, we calculate the distance between all our data points.”
- ☒ “Pick one random point, as initial point, and for the second point, instead of picking it randomly, we prioritize by assigning the probability of the distance.”

Correct! This one defines it and remember: The smarter initialization of K-mean clusters is called, K-means ++, and it helps to avoid getting stuck at these local optima. This is the default implementation of the K-means.

- ☐ “We start by having two centroids as far as possible between each other.”

3. What happens with our second cluster centroid when we use the probability formula?

1 point

- ☐ When we use the probability formula, we put less weight on the points that are far away. So, our second cluster centroid is likely going to be closer.
- ☐ When we use the probability formula, we put more weight on the points that are far away. So, our second cluster centroid is likely going to be more distant.
- ☒ When we use the probability formula, we put more weight on the lighter centroids, because it will take more computational power to draw our clusters. So, the second cluster centroid is likely going to be less distant.

Incorrect. This is not what happens when we use the probability formula.

- ☐ When we use the probability formula, we put less weight on the points that are far away. So, our second cluster centroid is likely going to be more distant.