K-Means Clustering

COSC 410: Applied Machine Learning
Fall 2025
Prof. Forrest Davis

September 4, 2025

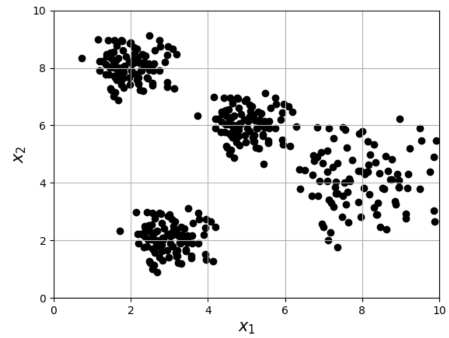

Figure 1: Sample data

Warm-up

1. Tell the person next to you about your favorite place to eat in town.
2. How many groups do you think are in Figure 2?

Logistics

- Lab 1 Due Friday
  - Submit one copy for a whole group and add your teammates as members to it
- Codelet 0 Due Friday September 5 at 11:59pm via Gradescop
- Codelet 1 released, please review it over the weekend

Learning Objectives

- Articulate the basic aims of k-means clustering
- Use Lloyd's algorithm to find clusters
- Distinguish different initialization strategies
- Articulate some core distinctions in training and evaluating models, especially around hyperparameter tuning

Summary: We layout the intuitions behind k-means clustering and give the standard algorithm used for fitting (Lloyd's Algorithm). We conclude with a discussion around initialization and broad approaches to hyperparameter tuning.

K-Means Basic Aims

K-MEANS CLUSTERING IS ONE APPROACH TO MODELING our intuitions from the warm-up. At its core, k-means clustering seeks to find k clusters in a dataset. The basic approach uses Lloyd's Algorithm.[1] More concretely, we are seeking k **centroids** which represent the point at the middle of the subset of the data belonging to a given cluster.

[1] The algorithm for k-means discussed here is also sometimes called Lloyd-Forgy method or simply the k-means algorithm due to its ubiquity.

Question

Where would you expect the centroids for Figure 2?

The approach is straightforward, we first select k initial centroids by sampling randomly k samples from our data to serve as centroids. Then, while not **converged**:

1. **Assignment:** Assign each point to the centroid closet to it
2. **Update:** Calculate new centroids based on the mean of the labeled points

Question

What do you think **convergence** means for k-means? How would you define closeness?

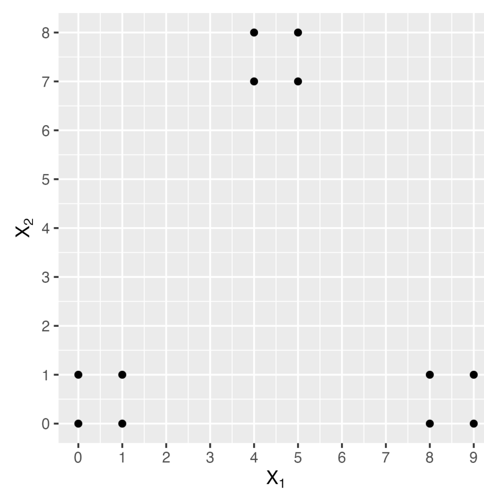

Figure 2: Sample data for group work

K-Means Simple Example

WITH YOUR SMALL GROUP WORK THROUGH the k-means algorithm with the following initial centroids and dataset (visualized in Figure 2).

$C_1^{(0)}$ is read as C one – centroid one – at time step 0.

Practice Problems

What are the final centroids and what cluster does each dataset belong to after apply k-means clustering with the initial centroids of:

$$C_1^{(0)} = (0,0) \text{ and } C_2^{(0)} = (9,1)$$

Point	$d(C_1)$	$d(C_2)$	Cluster	$d(C_1)$	$d(C_2)$	Cluster
(0,0)						
(0,1)						
(1,0)						
(1,1)						
(8,0)						
(8,1)						
(9,0)						
(9,1)						
(4,7)						
(5,7)						
(4,8)						
(5,8)						

Point	$d(C_1)$	$d(C_2)$	Cluster	$d(C_1)$	$d(C_2)$	Cluster
(0,0)						
(0,1)						
(1,0)						
(1,1)						
(8,0)						
(8,1)						
(9,0)						
(9,1)						
(4,7)						
(5,7)						
(4,8)						
(5,8)						

$$C_1^{(\ )} = (\ ,\ ) \text{ and } C_2^{(\ )} = (\ ,\ )$$

Practice Problems

What are the final centroids and what cluster does each dataset belong to after apply k-means clustering with the initial centroids of:

$$C_1^{(0)} = (0, 1) \text{ and } C_2^{(0)} = (1, 0)$$

Point	$d(C_1)$	$d(C_2)$	Cluster	$d(C_1)$	$d(C_2)$	Cluster
(0,0)						
(0,1)						
(1,0)						
(1,1)						
(8,0)						
(8,1)						
(9,0)						
(9,1)						
(4,7)						
(5,7)						
(4,8)						
(5,8)						

Point	$d(C_1)$	$d(C_2)$	Cluster	$d(C_1)$	$d(C_2)$	Cluster
(0,0)						
(0,1)						
(1,0)						
(1,1)						
(8,0)						
(8,1)						
(9,0)						
(9,1)						
(4,7)						
(5,7)						
(4,8)						
(5,8)						

$$C_1^{(\ )} = (\ ,\ ) \text{ and } C_2^{(\ )} = (\ ,\ )$$

K-Means++: Issues with Initialization

ABOVE, WE NOTED SOME ISSUES with a basic initialization strategy for k-means clustering. Let's dive into some deeper issues by considering the data plotted in Figure 3.[2]

[2] The example data here is adapted from this video

Figure 3: Dataset with four samples.

K-means++ Initialization:

1. Take one centroid $\mathbf{c}_1$, chosen uniformly at random from the dataset.
2. Take a new centroid $\mathbf{c}_i$, choosing a sample $\mathbf{x}_i$ with probability $\frac{D(\mathbf{x}_i)^2}{\sum_{j=1}^{m} D(\mathbf{x}_j)^2}$, where $D(\mathbf{x}_i)$ is the distance between the sample $\mathbf{x}_i$ and the closest centroid that was already chosen.
3. Repeat the previous step until all k centroids have been chosen.

Question

How does this new initialization effect model performance in Figure 3?

A Bit Deeper into K-Means

LET'S DIG A BIT DEEPER INTO what k-means optimizes.[3] In particular, we are seeking to optimize:

[3] The discussion here draws on Chapter 14 Section 3.6 from Hastie [2001].

$$\min_{C,\{m_k\}_1^K} \sum_{k=1}^{K} N_k \sum_{C(i)=k} \|x_i - m_k\|^2 \tag{1}$$

Lloyd's Algorithm Redux:

1. For a given cluster assignment C the distances between points and a centroid is minimized when the centroids are the means $\{m_1, \ldots, m_K\}$
2. Given a current set of means $\{m_1, \ldots, m_K\}$, the set of points which minimizes the distances in a cluster is obtained by labeling each point with the cluster mean closest to it That is,

$$C(i) = \arg\min_{1 \le k \le K} \|x_i - m_k\|^2.$$

3. Steps 1 and 2 are iterated

These facts guarantee that our algorithm will converge! However we may land at a local minimum that isn't the best globally.[4]

[4] Finding the truly optimal clusters for k as small as 2 is, in fact, NP-hard optimization problem. If you are interested see Dasgupta [2008].

General Considerations for Hyperparameter Tuning

There are no parameters in this model. The k we choose is a **hyperparameter**, it is not updated during training (i.e., it is not learned). There exist a number of techniques for hyperparameter tuning, some of which we will cover in a lab. We will take some time here, in the slides, to set up some basic terms that will be helpful later.

Before Next Class

- Read Chapter 6 of Hands on Machine Learning (linked on the course website) and complete the pre-class quiz

References

Sanjoy Dasgupta. The hardness of k-means clustering. *Technical Report CS2008-0916*, 2008.

Trevor Hastie. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, New York, 2001. ISBN 978-0-387-95284-0.