Study Guide: K-Means Clustering

Definition and Characteristics

K-Means Clustering is a type of unsupervised learning method that groups similar data points into clusters based on their features. It is a center-based clustering algorithm, which means that each cluster is defined by its center (also known as centroid) and each data point is assigned to the closest center.

Key Characteristics:

* Center-based clustering**: Each cluster is defined by its center (centroid)
* Assignment of data points: Each data point is assigned to the closest center
* Iterative process: The algorithm iterates until the clusters converge or a stopping criterion is met

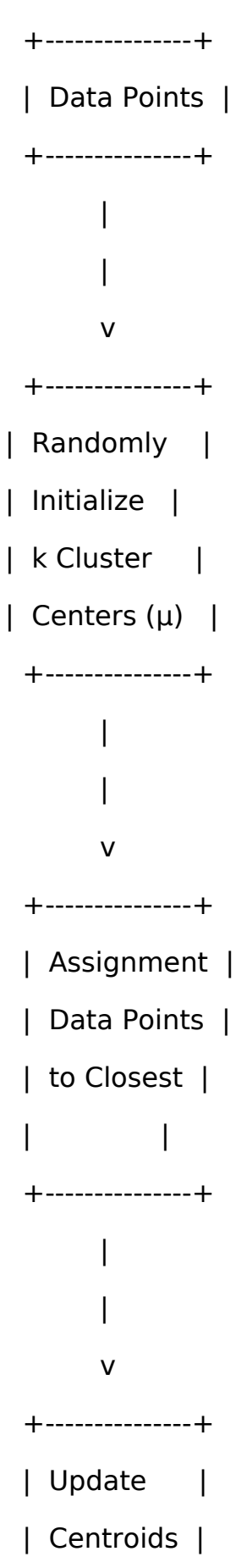
Applications

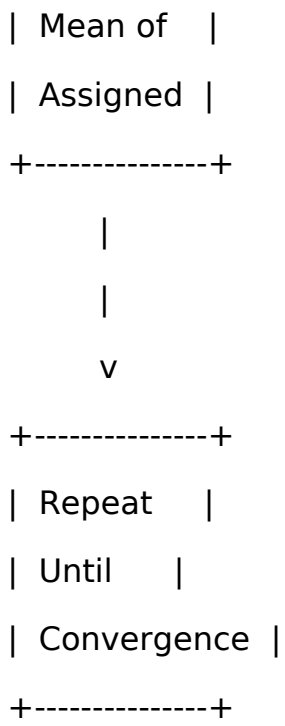
* Customer segmentation
* Image segmentation
* Gene expression analysis
* Recommendation systems

How K-Means Clustering Works

1. **Initialization**: Randomly initialize k cluster centers (centroids)
2. **Assignment**: Assign each data point to the closest centroid
3. **Update**: Update the centroid of each cluster by calculating the mean of all data points assigned to it
5. **Repeat**: Repeat steps 2-4 until convergence or a stopping criterion is met

Example Diagram (ASCII)





Theorem 7.4:

The projection of data points into the space spanned by the top k singular vectors of the data matrix brings data points closer to their cluster centers.

Summary of Key Points

* K-Means Clustering is a center-based clustering algorithm
* It iteratively assigns data points to the closest center and updates the center
* The algorithm converges when the clusters stop moving
* K-Means Clustering has applications in customer segmentation, image segmentation, gene expression analysis, and recommendation systems

Flashcards

Q1: What is K-Means Clustering?

K-Means Clustering is a type of unsupervised learning method that groups similar data points into clusters based on their features.

Q2: What is the key characteristic of center-based clustering?

Each cluster is defined by its center (centroid) and each data point is assigned to the closest center.

Q3: What is the iterative process in K-Means clustering?

The algorithm iterates until the clusters converge or a stopping criterion is met.

Q4: What is the application of K-Means Clustering in image segmentation?

It groups similar pixels into clusters, allowing for image segmentation.