Future Connect Media

Machine learning Part-D

Part of Future Connect Media's IT Course

By Abhishek Sharma

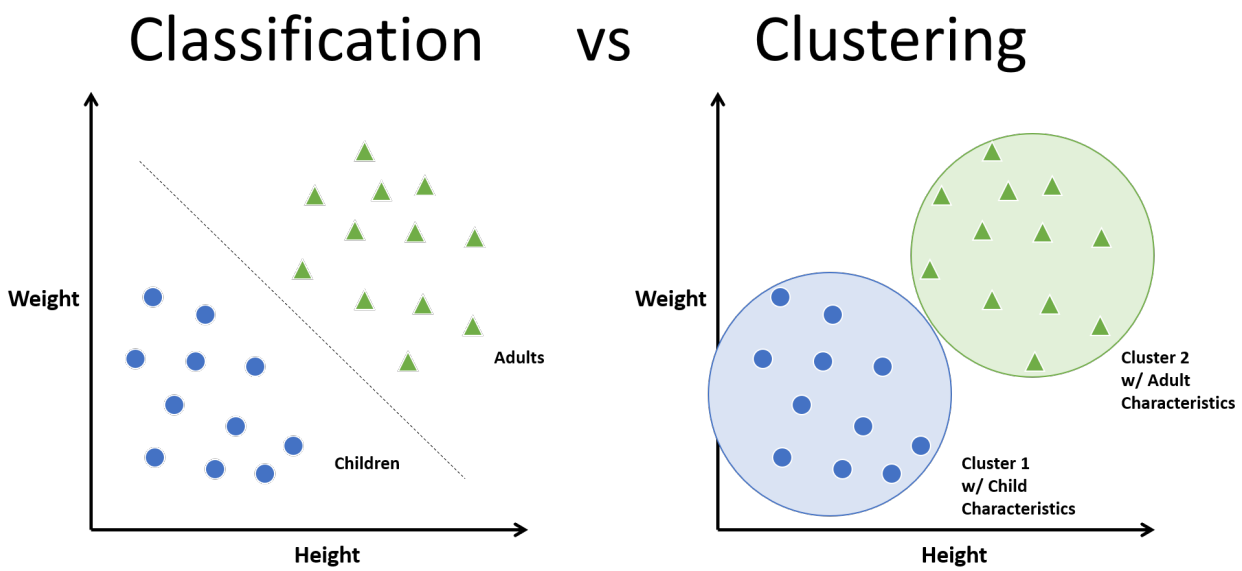
Future Connect Media

Topics to be covered:

- Clustering
- Model for Clustering
- K-Means Clustering
- Elbow Method

Future Connect Media

Clustering



- In machine learning too, we often group examples as a first step to understand a subject (data set) in a machine learning system. Grouping **unlabeled examples** is called **clustering**.
- As the examples are unlabeled, clustering relies on unsupervised machine learning. If the examples are labeled, then clustering becomes **classification**. For a more detailed discussion of supervised and unsupervised methods see

Future Connect Media

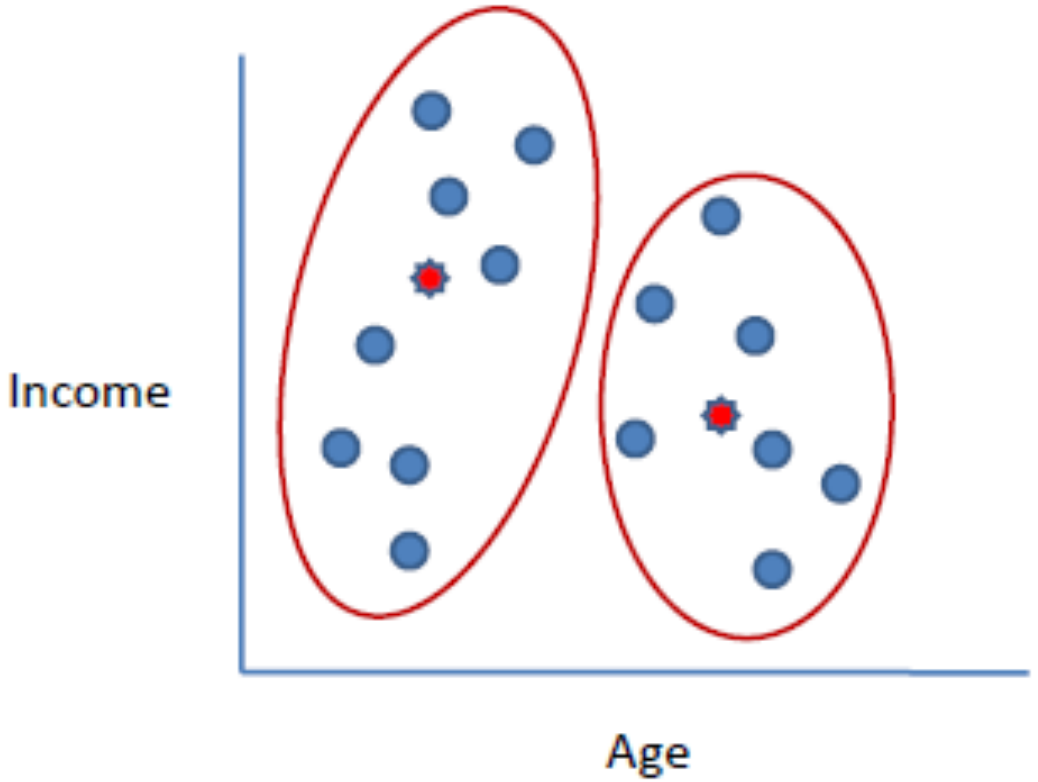
Models for Clustering

- Following are the models used for Clustering:
  - K-Means Clustering
  - Hierarchical Clustering

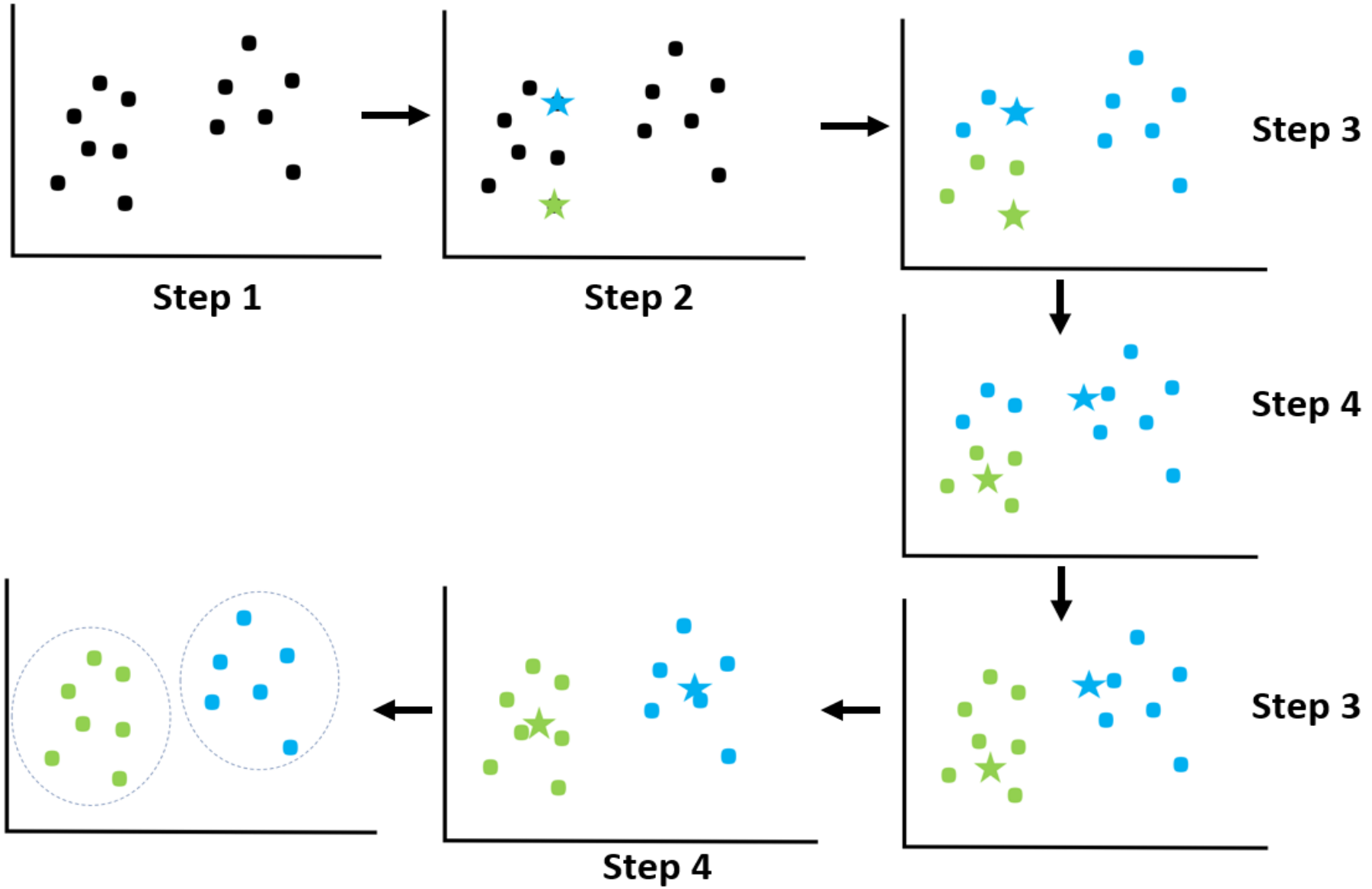
Future Connect Media

K-Means Clustering

- K-means clustering is one of the simplest and popular unsupervised machine learning algorithms.
- Typically, unsupervised algorithms make inferences from datasets using only input vectors without referring to known, or labelled, outcomes.
- The objective of K-means is simple: group similar data points together and discover underlying patterns. To achieve this objective, K-means looks for a fixed number (*k*) of clusters in a dataset.



Steps in K-Means Clustering



Future Connect Media

Elbow Method

- A fundamental step for any unsupervised algorithm is to determine the optimal number of clusters into which the data may be clustered. The **Elbow Method** is one of the most popular methods to determine this optimal value of k.