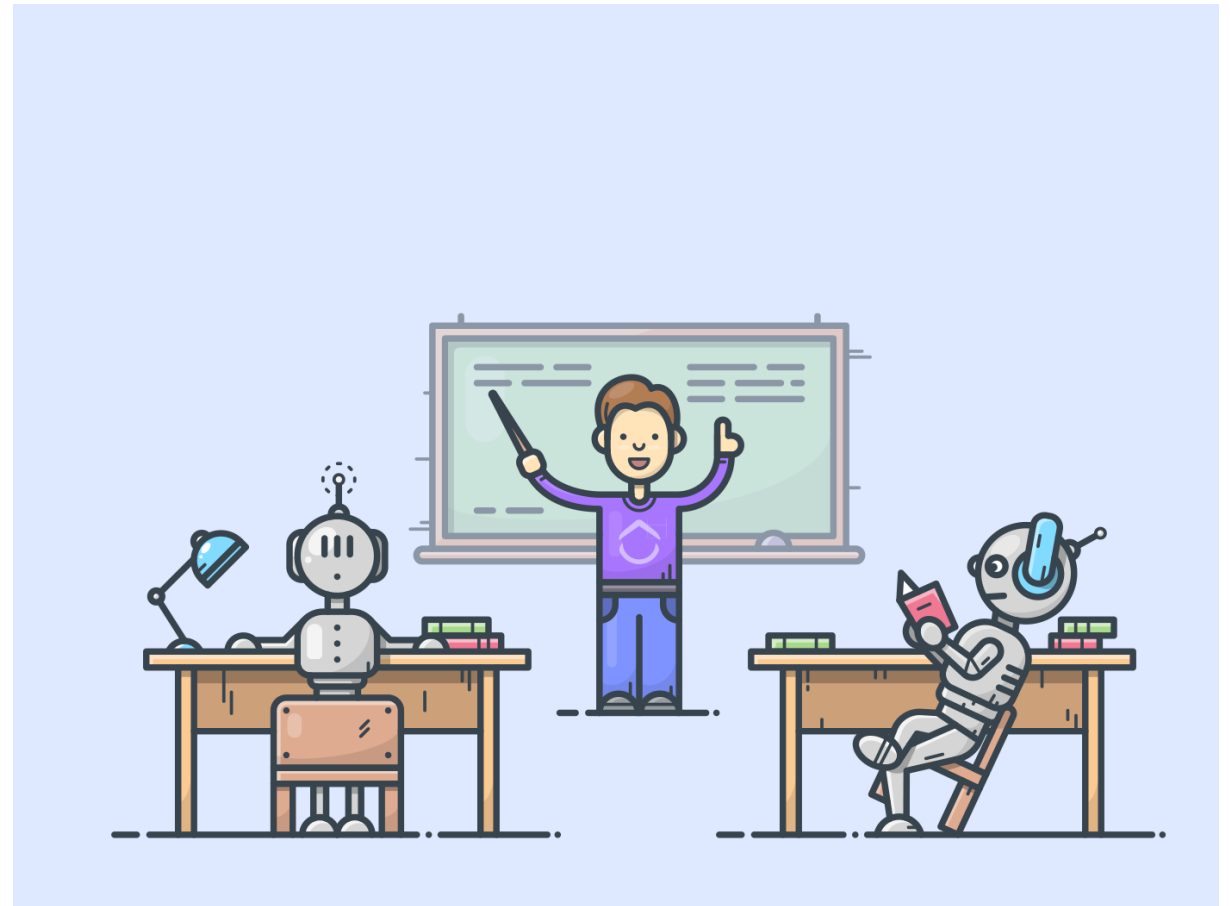


Unsupervised Learning



Objective

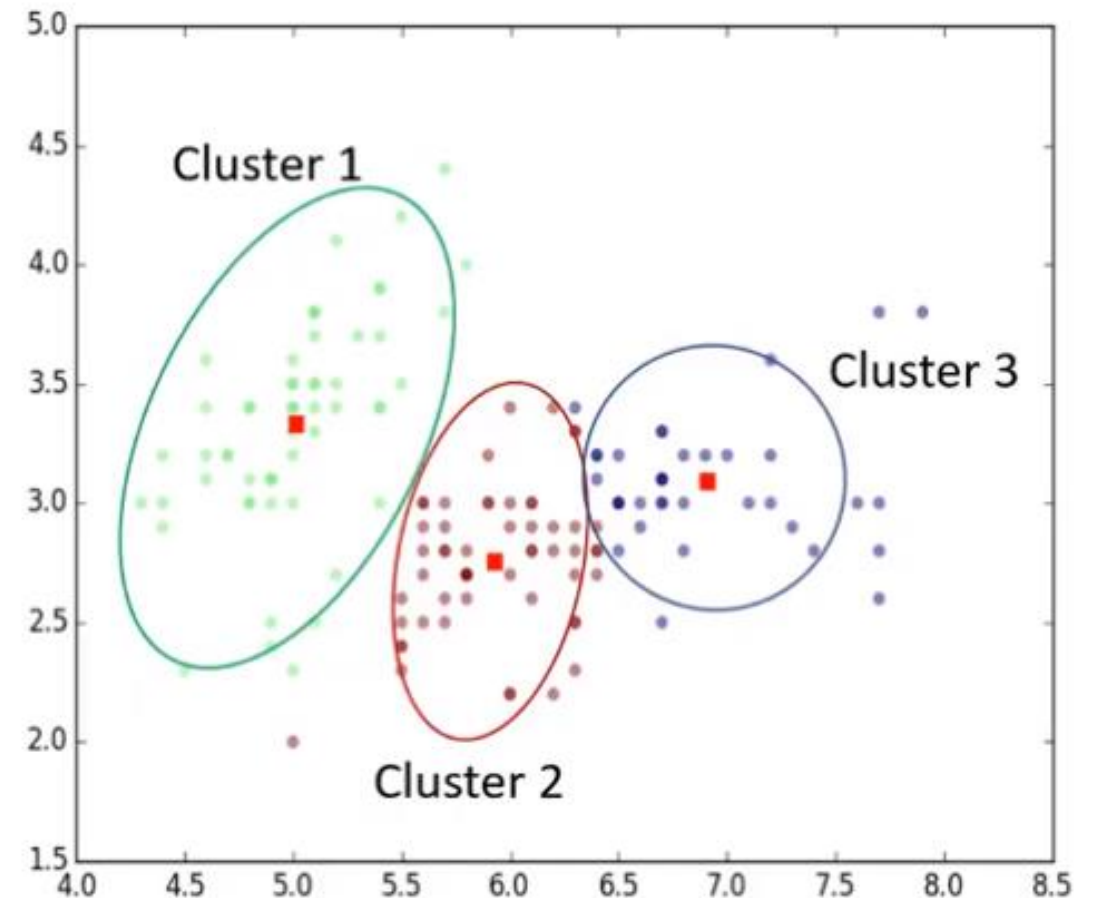
- Supervised learning vs. Unsupervised learning
- What is Clustering ?
- Case Study
- Clustering
- Similarity Assessment
- Inference
- Applications
- Where Clustering can be used ?

Supervised learning vs. Unsupervised learning

- Supervised learning: discover patterns in the data that relate data attributes with a target (class) attribute.
 - These patterns are then utilized to predict the values of the target attribute in future data instances.
- Unsupervised learning: The data have no target attribute.
 - We want to explore the data to find some intrinsic structures in them.

What is Clustering ?

- Clustering means finding clusters in a dataset, unsupervised.
- So, what is a cluster?
- A cluster is group of data points or objects in a dataset that are similar to other objects in the group, and dissimilar to data points in other clusters.



Case Study

- You have a customer dataset, and you need to apply customer segmentation on this historical data.
- Customer segmentation is the practice of partitioning a customer base into groups of individuals that have similar characteristics.
- Helps business to target specific groups of customers (ex: high profit & low risk customers or non-profile organization customers) by to more effectively allocating marketing resources.

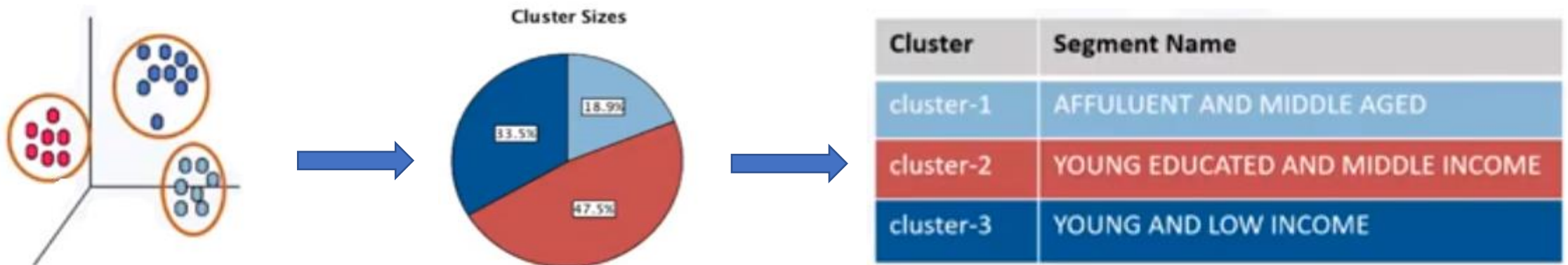
Customer Id	Age	Edu	Years Employed	Income	Card Debt	Other Debt	Address	DebtIncomeRatio	Defaulted
1	41	2	6	19	0.124	1.073	NBA001	6.3	0
2	47	1	26	100	4.582	8.218	NBA021	12.8	0
3	33	2	10	57	6.111	5.802	NBA013	20.9	1
4	29	2	4	19	0.681	0.516	NBA009	6.3	0
5	47	1	31	253	9.308	8.908	NBA008	7.2	0
6	40	1	23	81	0.998	7.831	NBA016	10.9	1
7	38	2	4	56	0.442	0.454	NBA013	1.6	0
8	42	3	0	64	0.279	3.945	NBA009	6.6	0
9	26	1	5	18	0.575	2.215	NBA006	15.5	1

Case Study...

- A general segmentation process is not usually feasible for large volumes of varied data.
- Therefore, you need an analytical approach to deriving segments and groups from large data sets.
- Customers can be grouped based on several factors: including age, gender, interests, spending habits, and so on.
- The important requirement is to use the available data to understand and identify how customers are similar to each other.
- One of the most adopted approaches that can be used for customer segmentation is clustering.

Clustering

- Clustering can group data only “unsupervised,” based on the similarity of customers to each other.
- It will partition your customers into mutually exclusive groups, ex: into 3 clusters.
- The customers in each cluster are similar to each other demographically.
- Now we can create a profile for each group, considering the common characteristics of each cluster.



Similarity Assessment

- Finally, we can assign each individual in our dataset to one of these groups or segments of customers.

Customer Id	Age	Edu	Years Employed	Income	Card Debt	Other Debt	Address	DebtIncomeRatio	Defaulted
1	41	2	6	19	0.124	1.073	NBA001	6.3	0
2	47	1	26	100	4.582	8.218	NBA021	12.8	0
3	33	2	10	57	6.111	5.802	NBA013	20.9	1
4	29	2	4	19	0.681	0.516	NBA009	6.3	0
5	47	1	31	253	9.308	8.908	NBA008	7.2	0
6	40	1	23	81	0.998	7.831	NBA016	10.9	1
7	38	2	4	56	0.442	0.454	NBA013	1.6	0
8	42	3	0	64	0.279	3.945	NBA009	6.6	0
9	26	1	5	18	0.575	2.215	NBA006	15.5	1

Customer ID	Segment
1	YOUNG AND LOW INCOME
2	AFFLUENT AND MIDDLE AGED
3	AFFLUENT AND MIDDLE AGED
4	YOUNG AND LOW INCOME
5	AFFLUENT AND MIDDLE AGED
6	AFFLUENT AND MIDDLE AGED
7	YOUNG AND LOW INCOME
8	YOUNG AND LOW INCOME
9	AFFLUENT AND MIDDLE AGED

Inference

- When we cross-join this segmented dataset, with the dataset of the product or services that customers purchase from our company.
- Would really help us to understand and predict the differences in individual customers' preferences and their buying behaviors across various products, allowing your company to provide highly personalized experiences for each segment.

Customer Id	Age	Edu	Years Employed	Income	Card Debt	Other Debt	Address	DebtIncomeRatio	Defaulted
1	41	2	6	19	0.124	1.073	NBA001	6.3	0
2	47	1	26	100	4.582	8.218	NBA021	12.8	0
3	33	2	10	57	6.111	5.802	NBA013	20.9	1
4	29	2	4	19	0.681	0.516	NBA009	6.3	0
5	47	1	31	253	9.308	8.908	NBA008	7.2	0
6	40	1	23	81	0.998	7.831	NBA016	10.9	1
7	38	2	4	56	0.442	0.454	NBA013	1.6	0
8	42	3	0	64	0.279	3.945	NBA009	6.6	0
9	26	1	5	18	0.575	2.215	NBA006	15.5	1

Customer ID	Segment
1	YOUNG AND LOW INCOME
2	AFFLUENT AND MIDDLE AGED
3	AFFLUENT AND MIDDLE AGED
4	YOUNG AND LOW INCOME
5	AFFLUENT AND MIDDLE AGED
6	AFFLUENT AND MIDDLE AGED
7	YOUNG AND LOW INCOME
8	YOUNG AND LOW INCOME
9	AFFLUENT AND MIDDLE AGED

Applications

- Customer segmentation is one of the popular usages of clustering.
- Cluster analysis also has many other applications in different domains.
- **Retail/Marketing**
 - Based on customer's demographic characteristics, identifying buying patterns of customer groups
 - Recommending new books or movies to customers
- **Banking**
 - Find clusters of normal transactions to find pattern of fraudulent credit card usage.
 - Identify clusters of customers to find loyal customers versus churn customers.

More Applications

- **Insurance**

- Analyse genuine claim clusters to identify false claim fraud.
- Evaluate insurance risk of customers based on their segments

- **Publication**

- Clustering is used to auto-categorize news based on its content,
- Tag news, then cluster it, so as to recommend similar news articles to readers

- **Medicine**

- Characterize patient behaviour, based on their similar characteristics, so as to identify successful medical therapies for different illnesses.

Where Clustering can be used ?

- Clustering can be used for one of the following purposes:
 - Exploratory data analysis
 - Summary generation or reducing the scale
 - Outlier detection, especially to be used for fraud detection
 - Noise removal
 - Finding duplicates in datasets
 - Pre-processing step for either prediction, other data mining tasks
 - Part of a complex system.



Thank You