Clustering

MKT 566

Instructor: Davide Proserpio

A few things…

- Survey
- Guest speaker next Monday

What is clustering?

- Clustering is an **unsupervised (machine) learning** technique
- Objective: **Group similar observations** together based on some of their characteristics
- Common applications:
  - Customer segmentation (e.g., by purchase behavior)
  - Grouping products by attributes
  - Grouping markets by demographic similarity

Why do we use clustering in marketing?

- Identify **distinct customer segments** with different behaviors/preferences.
- Tailor **targeted campaigns** and offers.
- Optimize **product positioning** and pricing.
- Improve **resource allocation**
  - Resources can then be shifted to growing clusters (e.g., emerging Gen Z customers)
  - Efficient Use of Marketing Budgets

Common clustering algorithms

- **K-Means Clustering**
  - Partitions data into *K* non-overlapping clusters
  - Fast and scalable
- **Hierarchical Clustering**
  - Builds a tree of clusters (dendrogram)
  - Useful for exploratory analysis
- **DBSCAN (Density-Based)**
  - Detects clusters of varying shape and size
  - Handles noise well
  - Well-suited for geographical data

Common clustering algorithms

- **K-Means Clustering**
  - Partitions data into *K* non-overlapping clusters
  - Fast and scalable
- **Hierarchical Clustering**
  - Builds a tree of clusters (dendrogram)
  - Useful for exploratory analysis
- **DBSCAN (Density-Based)**
  - Detects clusters of varying shape and size
  - Handles noise well
  - Well-suited for geographical data

Example: Customer segmentation

- Variables: Recency (first time they bought), Frequency (how often they buy), Spend
- Apply a clustering algorithm
- Result:
  - Cluster 1: High spenders, frequent buyers
  - Cluster 2: Low spenders, infrequent buyers
  - Cluster 3: New customers
  - Cluster 4: Lapsed customers

Visualizing clusters

PROBLEM: You often have more than two variables, so how can you plot the data?

PCA (Principal Component Analysis): a technique used for dimensionality reduction

- From N dimensions to two → very useful for visualization!
- Very helpful for creating perceptual maps

PCA in a nutshell

Suppose you are analyzing customer survey data with **50 questions**

- Some questions overlap (“How satisfied are you with service?” vs. “Would you recommend us?”).

Challenge: How do we simplify the data without losing too much information?

- PCA looks for the **underlying dimensions** that explain the most variation in the data.
- Instead of 50 survey questions, maybe there are just **2 main themes**:
  - **“Overall satisfaction”** (combining many service-related questions)
  - **“Perceived value”** (combining many price/benefit-related questions)
- These become **principal components**.

PCA vs. Clustering

Aspect	Clustering	PCA
Purpose	Group similar points	Reduce dimensionality / visualize
Output	Discrete cluster labels	Continuous components
Focus	Segments of *customers/items*	Relationships among *variables*
Uses	Market segmentation, recommender systems	Visualization, noise reduction, feature extraction
Analogy	"Which students form study groups?"	"Which directions explain most variance in exam scores?"

PCA + Clustering

- **Clustering without PCA** = labels only, no intuition about *what dimensions matter*.
- **PCA without clustering** = map only, but no hard group definitions.
- **Both together** = the best of both worlds:
  - PCA provides the *map* (market structure).
  - Clustering provides the *segments* (actionable groups).

Clustering best practice

- **Standardize** numerical variables before clustering
  - Not needed if variables use the same scale (but often is not the case)
- Try multiple algorithms and compare
- Visualizing clustering often helps
  - Check if clusters are **well-separated** or **overlapping**

Some high-level details on k-means

1. **Choose K** (number of clusters you want).
2. Randomly place **K “centroids”** in the data space.
3. Repeat until centroids don’t move (or max iterations reached):
   - **Assign** each data point to the nearest centroid.
   - **Update** each centroid’s position to the mean of its assigned points.

Choosing the number of clusters K

- Too few clusters → **lose detail**
- Too many clusters → **overfit**
  - Example: if you set K= number of points, each point becomes its own cluster. The fit is perfect, but you learn nothing!
- Common methods:
  - **Elbow method:** Look for the point where adding more clusters gives diminishing returns in **variance explained**
    - Plot the **sum of squared distances from each point to its cluster center**, across all clusters.

PCA example

Let’s use a dataset from R for Marketing Students

- A survey in which respondents were asked to rate four brands of office equipment on six dimensions.

	brand	large_choice	low_prices	service_quality	product_quality	convenience	preference_score
	<char>	<num>	<num>	<num>	<num>	<num>	<int>
1:	OfficeStar	5.2	2.1	4.2	3.7	2.7	5
2:	PaperNCo	4.4	4.5	2.3	2.6	1.4	3
3:	OfficeEquipment	3.9	2.6	3.1	3.1	4.7	3
4:	Supermarket	2.3	4.1	1.8	2.9	5.1	1

Clustering (+PCA) example using k-means

- Let's use a dataset from R for Marketing Students
  - A survey in which 40 respondents were asked to rate the importance of several store attributes when buying equipment

	respondent_id	variety_of_choice	electronics	furniture	quality_of_service	low_prices	return_policy	professional	income	age
	<num>	<num>	<num>	<num>	<num>	<num>	<num>	<num>	<num>	<num>
1:	1	8	6	6	3	2	2	1	40	45
2:	2	6	3	1	4	7	8	0	20	41
3:	3	6	1	2	4	9	6	0	20	31
4:	4	8	3	3	4	8	7	1	30	37
5:	5	4	6	3	9	2	5	1	45	56
6:	6	8	4	3	5	10	6	1	35	28