Unsupervised learning

Training head: Koby

31-10-23

Unsupervised learning

- Mostly clustering tasks

Supervised vs. Unsupervised Learning

data iku
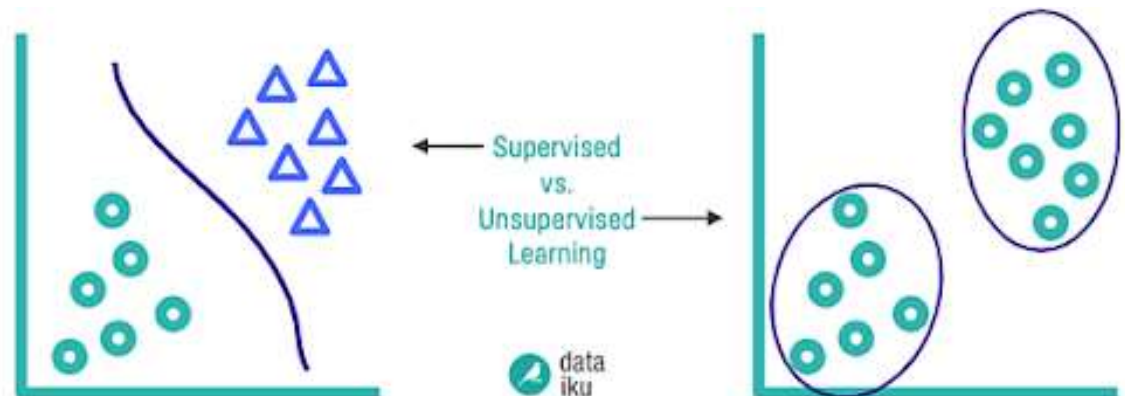
K-means clustering

- Partition the objects into K clusters such that the sum of squared distances between the objects and their assigned cluster mean is minimized

- A further application of K-means, hierarchical clustering, creates a hierarchical tree-like structure to represent the relationships between the clusters.

Choosing the right k value

Elbow plot:

- The number where inertia (any evaluation criteria from the dataset) decreases the least (becomes almost constant) is chosen as k
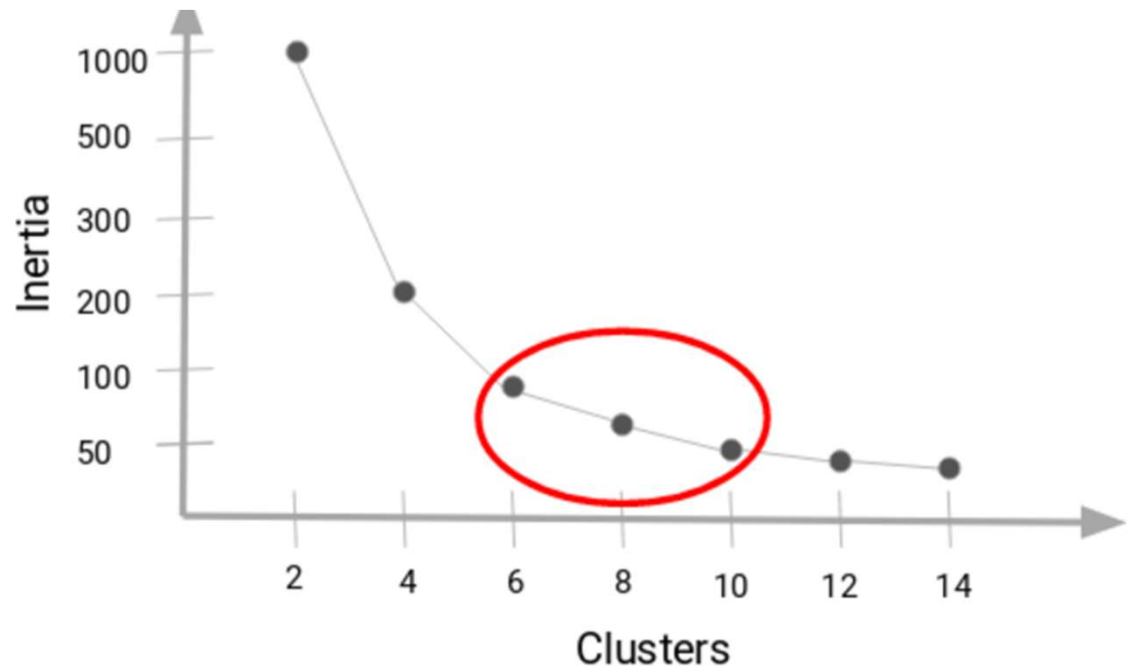
K-means clustering in Fintech

e.g. Bank wants to give credit card promotions to customers of 9 defined clusters consisting of 3 income levels (low, average and high), and 3 credit ratings (weak, average and strong credit worthiness)
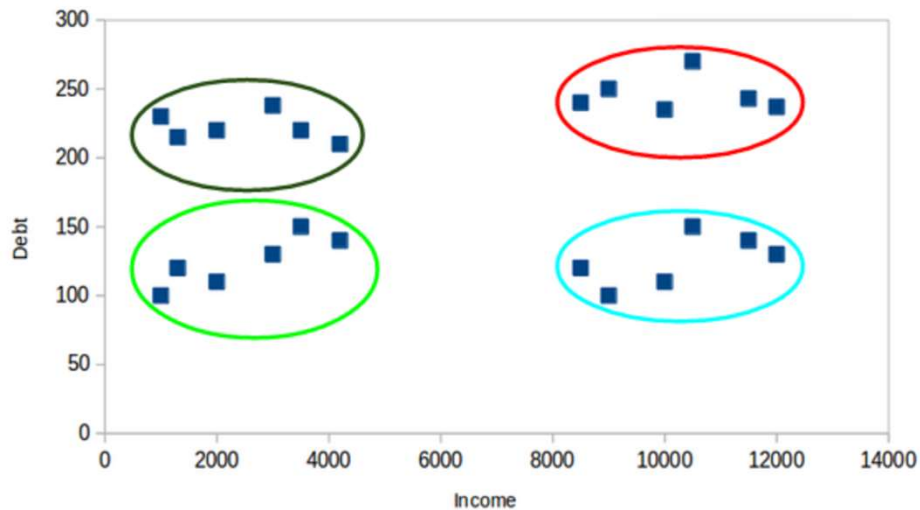
Credit card promotion example

- Using debt as the single indicator of credit rating [in reality, results from a credit rating neural network can be provided so customers can be assessed holistically]

- Only 4 clusters [more clusters could reveal subtler patterns e.g. which shops would lower middle income patronise more often than the other income groups, and how does this differ from upper middle?]

- There could be more than 2 factors that customers can be assessed by, other than income and credit worthiness, spending level at recommended shops could be investigated. 3 factors may require PCA [https://en.wikipedia.org/wiki/Principal_component_analysis]

Process of clustering

1. Select the value of k
2. Choose k random points from the dataset as the starting centroids
3. Assign all points to their current closest centroid (with Euclidean distance for k-means or argmax in EMA https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm)
4. Calculate the new centroids (this differs from the starting centroids)
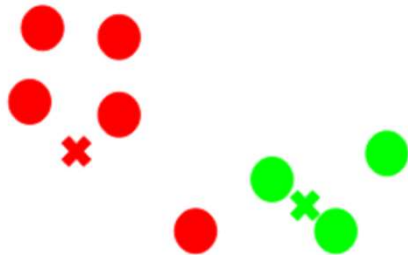
Process of clustering cont.

- Repeat steps 3 and 4 until a stopping criteria is met
- Stopping criteria: 1) Centroids of newly formed clusters do not change 2) Points remain in the same cluster   3) Maximum number of iterations is reached

Applications of clustering (non-fintech)

- Image segmentation (an object's pixels should be similar and similar pixels should be grouped together)

- Recommendation engines (in Spotify, e-commerce, Youtube, Instagram, what posts/videos/products did similar people like or buy)

K-means clustering example

- https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/
- PCA for k-means: https://365datascience.com/tutorials/python-tutorials/pca-k-means/
- Also search for hierarchical clustering as a more computationally complex algorithm.

Project 2 Q+A

Stay in Touch!

- My Telegram: @KC
- My school email: e0960154@u.nus.edu