State of the art of unsupervised learning and where it is being currently used

Yue Liu

What is unsupervised learning?

In supervised learning, we need to get a collection of labelled target data, or at least some scoring system. However, in unsupervised learning, we don't have a specific goal. Unsupervised learning can be considered as a way to reduce the dimensions. Most of the unsupervised learning are a form of cluster analysis: some objects in the clusters are very similar while others are distinct.

Where to use unsupervised learning?

It depends, we can even apply classification in the un-supervised learning, as long as the goal is to identify similarities between the input. For example, we can use k-Nearest Neighbors (k-nn) to classify the similar groups of inputs, if we did not pick up an output label.

K-means clustering method is a frequently used method, which contains four steps: (1) select k "classifier" points at random; (2) classify according to closest classifier point; (3) Replacing the classifier points by the centroids; (4) Repeating steps 2 and 3 until set membership stabilizes.

Where does unsupervised learning currently used?

In the real world, we can use k-means clustering to site cell phone towers; we can also use Fuzzy c-Means Clustering to analyze Gene Expression Data. Using self-organizing map, neural network based clustering can transform a dataset into a topology-preserving 2D map. Dimension reduction can be used to lower the complexity, and it will be good to preprocessing the data for supervised learning.

References

[1] Marsland, S. (2014). *Machine learning: an algorithmic perspective*. Chapman and Hall/CRC.
[2] Ratne, Alex., Bach, Stephen., Varma, Paroma., Ré, Chris (2017-07-16). An Overview of Weak Supervision [Web blog post]. Retrieved from https://www.snorkel.org/blog/weak-supervisionL.
[3] Mastering Machine Learning: A Step-by-Step Guide with MATLAB
Retrieved from https://www.mathworks.com/discovery/unsupervised-learning.html.