

Research Notes-Unsupervised learning

1. What is clustering used for?

When you're doing unsupervised learning and have a huge data set of unlabeled data that you want to obtain knowledge from, Clustering is used to create classes based on the statistical similarity of each object. In other words clustering groups sets of data based on similarities.

2. To what previous content in the course can you relate clustering and why?

ID3 decision trees relate to clustering because it involves grouping subsets of data based on a specific similarity among all data.

3. Search in a formal source (book, research article, or a certified online course) the concept of similarity in Artificial intelligence.

Similarity is an area which goals are to determine how similar or related objects are. The four common methods for similarity learning and metric distance learning are:

- Regression
 - Classification
 - Ranking
 - Locality
- [Wikipedia... mehhh]

A theoretical approach to similarity is that objects are represented as collections of features, and similarity is described as a feature-matching process. Similarity serves as an organizing principle by which individuals classify objects, form concepts and make generalizations. It is employed to explain errors in memory and pattern recognition. The theoretical analysis of similarity has been dominated by geometric models. Most analyses of proximity data have been metric in nature; they assume objects can be adequately represented as points in some coordinate space. It argued that dimensional representations are appropriate for certain stimuli, but it is better to represent faces, countries, or personalities in terms of many qualitative features rather than in terms of a few quantitative dimensions. [Features of similarity, Amos Tversky]

-Features of Similarity, Amos Tversky, Hebrew University
[<http://www.ai.mit.edu/projects/dm/Tversky-features.pdf>]