



Model-Based Clustering Methods

- Attempt to optimize the fit between the data and some mathematical model
- Statistical and AI approach
 - Conceptual clustering
 - A form of clustering in machine learning
 - Produces a classification scheme for a set of unlabeled objects
 - Finds characteristic description for each concept (class)
 - COBWEB (Fisher'87)
 - A popular a simple method of incremental conceptual learning
 - Creates a hierarchical clustering in the form of a **classification tree**
 - Each node refers to a concept and contains a probabilistic description of that concept

March 26, 2013

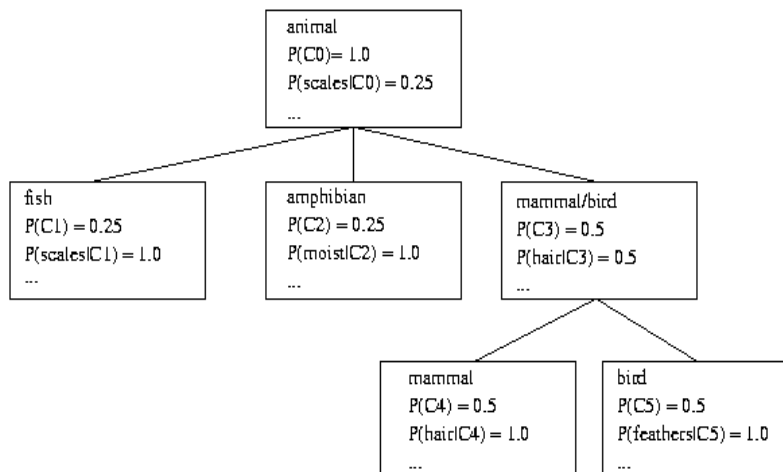
Data Mining: Concepts and Techniques

1



COBWEB Clustering Method

A classification tree



March 26, 2013

Data Mining: Concepts and Techniques

2



Cobweb

- Category utility uses conditional probabilities to measure
 - Intraclass similarity
 - Interclass dissimilarity
- Category utility comparisons used to
 - Assign instances to classes
 - Create new classes
 - User need not specify how many classes



Clustering: COBWEB

- Start with a tree consisting of an empty node
- Add instances one at a time, updating the tree after each addition
 - Update by finding the correct leaf in which to place the new instance
 - Placing the new instance can require restructuring the tree
- Base updating decisions on
 - **partition utility**
 - **category utility**