Which of these terms describes the first major task in the data classification process?

- ○ Choose training data
- ○ Classify
- ◉ Learning
- ○ Analyze the training data for possible "noise" in the database

Which of these is typically used to represent the sought-after mapping function in the first step of the classification process?

- ○ Decision Trees
- ○ Mathematical formula
- ○ Classification rules
- ◉ All of these answer choices are correct.

In BBNs, each arc in the acyclic graph represents:

- ○ An educated guess as to a causal relationship between the two connected nodes
- ○ None of these answer choices are correct.
- ◉ The probabilistic dependency of the two connected nodes
- ○ A non-dependent relationship of the two connected nodes

This type of classification builds a classifier using both labeled and unlabeled data.

- ◉ Semi-supervised classification
- ○ Multi-class classification
- ○ Instance-based
- ○ Support vectors

Clustering is an approach that uses which of these learning techniques?

- ○ Examples
- ○ Listening
- ◉ Observation
- ○ All of these answer choices are correct.

If the classes of the training data are an unknown, we can apply which of these computational algorithms to attempt to find useable classes?

- ○ Pruning Analysis
- ◉ Clustering
- ○ Binary Sort
- ○ Bayesian Analysis

Mining time-series data often involves:

- ○ Exactly matching sub-sequences of the time-series against a similar preselected subsequence
- ○ Analyzing random sub-sequences from the series for known patterns
- ○ None of these answer choices are correct.
- ◉ Searching for data sequences in the series that are similar to a preselected "query" sequence

✓ **ANSWER:**

Which of these situations would be considered anomaly detection in a semi-supervised mode?

- ○ No training set exists with even class labels for normal or anomalous objects
- ○ None of these answer choices are correct.
- ○ Training set exists with full information needed about both normal and anomalous objects
- ◉ Training set exists with information needed for normal objects only

Hierarchical clustering methods:

- ○ All of these answer choices are correct.
- ○ Are always top-down in their approach to forming hierarchies
- ○ Are impervious to the choice of split/merge points
- ◉ Start with individual objects as clusters or the entire data set as a single cluster

The most useful feature of CLIQUE is that:

- ○ It is based on an already familiar algorithm, so its properties are well understood
- ○ It can summarize the list of cells that comprise a cluster with a small set of inequalities
- ○ None of these answer choices are correct.
- ◉ It finds clusters in subspaces efficiently

Which of these is not typically thought to be an application area where outlier or anomaly detection is paramount?

- ◉ Detection of books of interest to a library client
- ○ Ecosystem disturbances
- ○ Fraud detection
- ○ Intrusion detection

If we find the same four five-second sub-sequences occur, in order, every 45 seconds in a radio signal, this is an example of a:

- ○ Trend movement over time
- ◉ Cyclic movement over time
- ○ Seasonal variation movement over time
- ○ Random movement over time