

The background of the slide is a complex, abstract composition. It features a network of thin, reddish-brown lines forming a web-like structure. Scattered throughout are numerous small, green and blue dots, resembling data points or particles. On the left side, there is a vertical strip with a grid of small, light-colored squares. In the center, a white, angular shape, resembling a stylized 'V' or a folded piece of paper, serves as a backdrop for the title. The overall color palette is muted, with earthy tones and a touch of digital vibrancy from the dots and lines.

What Is Pattern Discovery? Why Is It Important?

What Is Pattern Discovery?

□ What are patterns?

□ **Patterns**: A set of items, subsequences, or substructures that occur frequently together (or ^{随后的} strongly correlated ^{子结构}) in a data set

□ Patterns represent **intrinsic** and **important properties** of datasets

□ **Pattern discovery**: Uncovering patterns from massive data sets

□ Motivation examples:

- What products were often purchased together?
- What are the subsequent purchases after buying an iPad?
- What code segments likely contain copy-and-paste bugs?
- What word sequences likely form phrases in this corpus?

Pattern Discovery: Why Is It Important?

- ❑ Finding **inherent regularities** in a data set
- ❑ **Foundation** for many essential data mining tasks
 - ❑ Association, correlation, and causality analysis
 - ❑ Mining sequential, structural (e.g., sub-graph) patterns
 - ❑ Pattern analysis in spatiotemporal, multimedia, time-series, and stream data
 - ❑ Classification: Discriminative pattern-based analysis
 - ❑ Cluster analysis: Pattern-based subspace clustering
- ❑ Broad applications
 - ❑ Market basket analysis, cross-marketing, catalog design, sale campaign analysis, Web log analysis, biological sequence analysis