

The background of the slide is a complex, abstract composition. It features a dark, reddish-brown base with a network of thin, light-colored lines forming a web-like structure. Overlaid on this are various data visualizations: a grid of small, light-colored plus signs, a series of small, colorful dots (green, blue, yellow) connected by lines, and a large, semi-transparent white triangle that frames the central text. In the bottom left corner, there is a small, rectangular inset showing a cluster of orange and red dots with a white grid overlay.

Lecture 2. Pattern Discovery: Basic Concepts

Lecture 2. Pattern Discovery: Basic Concepts

- ❑ What Is Pattern Discovery? Why Is It Important?
- ❑ Basic Concepts: Frequent Patterns and Association Rules
- ❑ Expressing Patterns in Compressed Forms: Closed Patterns
vs. Max-Patterns



Session 1: 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 strongly correlated) 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