

The background of the slide is a complex, abstract composition. It features a network of thin, light-colored lines forming a web-like structure. Overlaid on this are various data points: small green dots, larger orange and red dots, and some blue dots. There are also faint, larger-scale geometric shapes and patterns in shades of purple, blue, and orange. The overall aesthetic is technical and data-driven.

# **Lecture 6. Constraint-Based Pattern Mining**

# Lecture 6. Constraint-Based Pattern Mining

---

- ❑ Why Constraint-Based Mining?
- ❑ Different Kinds of Constraints: Different Pruning Strategies
- ❑ Constrained Mining with Pattern Anti-Monotonicity
- ❑ Constrained Mining with Pattern Monotonicity
- ❑ Constrained Mining with Data Anti-Monotonicity
- ❑ Constrained Mining with Succinct Constraints
- ❑ Constrained Mining with Convertible Constraints
- ❑ Handling Multiple Constraints



The background of the slide is a complex, abstract composition. It features a network of thin, light-colored lines forming a web-like structure. Overlaid on this are various data visualization elements: a grid of small grey plus signs, clusters of green and blue dots, and a prominent orange and red cluster on the left side. The overall color palette is muted, with shades of brown, grey, and white, accented by the colors of the data points.

# **Session 1. Why Constraint-Based Mining?**

# Why Constraint-Based Mining?

---

- ❑ Finding **all** the patterns in a dataset **autonomously**? — unrealistic!
  - ❑ Too many patterns but not necessarily user-interested!
- ❑ Pattern mining should be an **interactive** process
  - ❑ User directs what to be mined using a **data mining query language** (or a graphical user interface)
- ❑ Constraint-based mining
  - ❑ User flexibility: provides **constraints** on what to be mined
  - ❑ Optimization: explores such constraints for efficient mining
    - ❑ **Constraint-based mining**: Constraint-pushing, similar to push selection first in DB query processing

# Constraints in General Data Mining

A data mining query can be in the form of a meta-rule or with the following language primitives

- **Knowledge type constraint:**
  - Ex.: classification, association, clustering, outlier finding, ....
- **Data constraint** — using SQL-like queries
  - Ex.: find products sold together in NY stores this year
- **Dimension/level constraint**
  - Ex.: in relevance to region, price, brand, customer category
- **Rule (or pattern) constraint**
  - Ex.: small sales (price < \$10) triggers big sales (sum > \$200)
- **Interestingness constraint**
  - Ex.: strong rules:  $\text{min\_sup} \geq 0.02$ ,  $\text{min\_conf} \geq 0.6$ ,  $\text{min\_correlation} \geq 0.7$

# Meta-Rule Guided Mining

---

- ❑ A meta-rule can contain partially instantiated predicates & constants
  - ❑  $P_1(X, Y) \wedge P_2(X, W) \Rightarrow \text{buys}(X, \text{"iPad"})$
- ❑ The resulting mined rule can be
  - ❑  $\text{age}(X, \text{"15-25"}) \wedge \text{profession}(X, \text{"student"}) \Rightarrow \text{buys}(X, \text{"iPad"})$
- ❑ In general, (meta) rules can be in the form of
  - ❑  $P_1 \wedge P_2 \wedge \dots \wedge P_l \Rightarrow Q_1 \wedge Q_2 \wedge \dots \wedge Q_r$
- ❑ Method to find meta-rules
  - ❑ Find frequent ( $l + r$ ) predicates (based on *min-support*)
  - ❑ Push constants deeply when possible into the mining process
    - ❑ Using constraint-push techniques introduced in this lecture
  - ❑ Also, push `min_conf`, `min_correlation`, and other measures as early as possible (measures acting as constraints)