

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 large, semi-transparent white trapezoidal shape that serves as a backdrop for the title. In the bottom-left corner, there is a small, square inset image showing a dense cluster of orange and red dots with a horizontal band of pink and white squares.

# **Session 3: Mining Quantitative Associations**

# Mining Quantitative Associations

- ❑ Mining associations with numerical attributes

- ❑ Ex.: Numerical attributes: **age** and **salary**

- ❑ Methods

- ❑ Static discretization based on predefined concept hierarchies

- ❑ Data cube-based aggregation

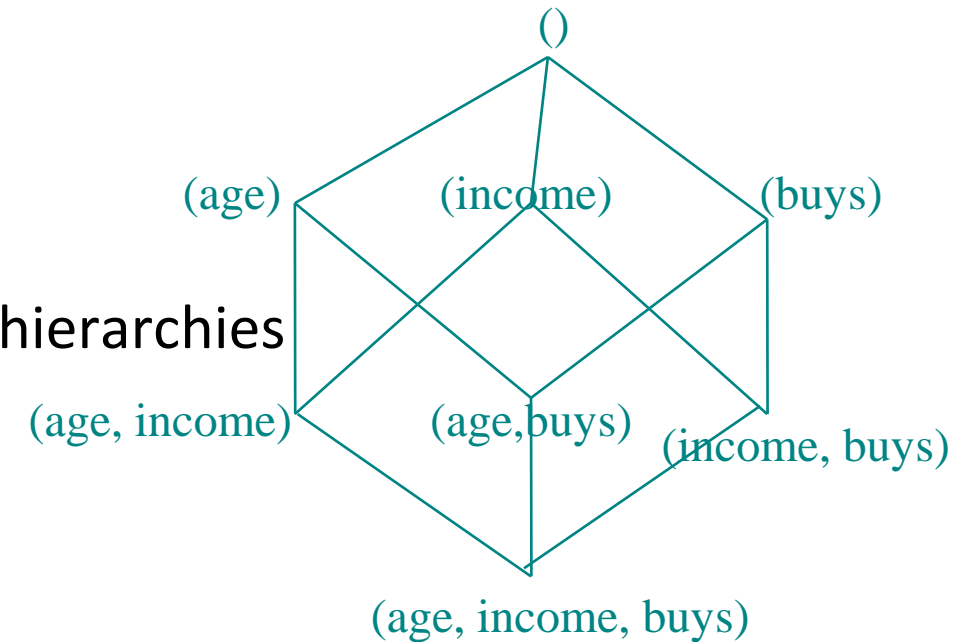
- ❑ Dynamic discretization based on data distribution

- ❑ Clustering: Distance-based association

- ❑ First one-dimensional clustering, then association

- ❑ Deviation analysis:

- ❑ Gender = female  $\Rightarrow$  Wage: mean=\$7/hr (overall mean = \$9)



# Mining Extraordinary Phenomena in Quantitative Association Mining

---

- ❑ Mining extraordinary (i.e., interesting) phenomena
  - ❑ Ex.: Gender = female  $\Rightarrow$  Wage: mean=\$7/hr (overall mean = \$9)
  - ❑ LHS: a subset of the population
  - ❑ RHS: an extraordinary behavior of this subset
- ❑ The rule is accepted only if a statistical test (e.g., Z-test) confirms the inference with high confidence
- ❑ Subrule: Highlights the extraordinary behavior of a subset of the population of the super rule
  - ❑ Ex.: (Gender = female)  $\wedge$  (South = yes)  $\Rightarrow$  mean wage = \$6.3/hr
- ❑ Rule condition can be categorical or numerical (quantitative rules)
  - ❑ Ex.: Education in [14-18] (yrs)  $\Rightarrow$  mean wage = \$11.64/hr
- ❑ Efficient methods have been developed for mining such extraordinary rules (e.g., Aumann and Lindell@KDD'99)