

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 circular dots. On the left side, there is a vertical strip containing a series of small, light-colored squares, some of which are highlighted in orange. The overall color palette is muted, with earthy tones and a soft, hazy atmosphere.

# **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

年龄可以每隔5年,薪水可以每隔100\$

改进.

□ Data cube-based aggregation

缺点: data distribution 可以随数据集的不同而不同.

□ Dynamic discretization based on data distribution

□ Clustering: Distance-based association

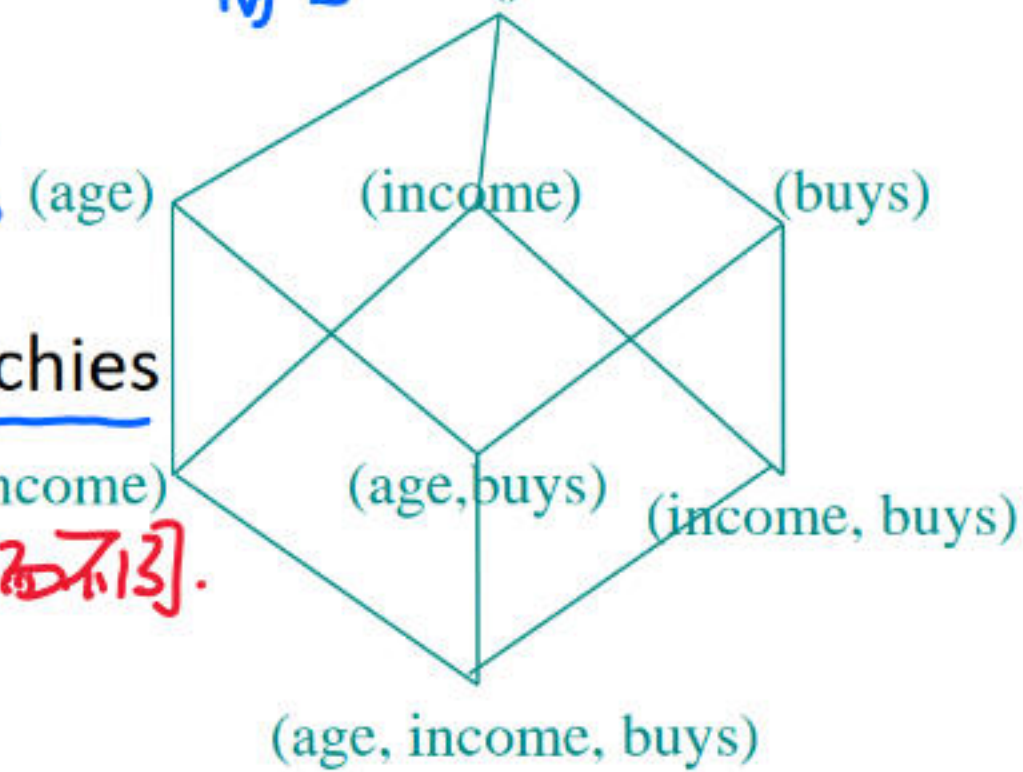
□ First one-dimensional clustering, then association

□ Deviation analysis:

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

deviation, 有了偏差

static discretization  
静态离散化.





# 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 → 增加LHS
  - Ex.: (Gender = female)  $\wedge$  (South = yes)  $\Rightarrow$  mean wage = \$6.3/hr
- Rule condition can be categorical or numerical (quantitative rules) 定量规则
  - 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)