Previous Class:

Relational Database
Data Warehouse
Transactional Database

Other Kinds of Data:

Time-Related or Sequence Data
Data Stream
Spatial Data
Engineering Design Data
Hypertext and Multimedia Data
Web Data

Data Mining Functionalities

1. **Class / Concept Description**
2. **Mining Frequent Patterns, Associations, and Correlations**
3. **Classification and Regression for Predictive Analysis**
4. **Cluster Analysis**
5. **Outlier Analysis**

(Kinds of Patterns that can be mined)

1. **Class / Concept Description**

   "Data Entries that can be associated either with classes or concepts"

   Ex: Classes of items for the sale (Class)

   Ex: Concepts of customers include budget spenders and big spenders

   How to derive these descriptions?

   1. Data Characterization – target class
      (Summarization for the data of the class under study)
   2. Data Discrimination – target class and contrasting class
      (Comparison of the target class with one or more number of contrasting classes)

4

1. Class / Concept Description (Contd..)

The output of the data is presented in various forms, include, pie charts, bar charts, curves, multi-dimensional data cubes, tables, cross tab

Ex:
1. Summarize the characteristics of customers who spend more than $5000 a year at AllElectronics.

2. A customer relationship manager at AllElectronics may want to compare two groups of customers—those who shop for computer products **regularly** (e.g., more than twice a month) and those **who rarely shop** for such products (e.g., less than three times a year).

Data Mining Functionalities

2. Mining Frequent Patterns, Associations, and Correlations

Frequent Patterns : Patterns that occur frequently

(Bread and Milk), (Milk and Egg), (Bread, Milk and Egg)
            14   11   7

Frequent Itemsets: Set of Items that occur frequently

"**Mining frequent patterns that leads to discovery of interesting associations and correlations within the data**"

6

2. Mining Frequent Patterns, Associations, and Correlations (Contd..)

Single dimensional association rule:
Association rule that contains a single predicate with support and confidence

$$buys(X, \text{``computer''}) \Rightarrow buys(X, \text{``software''}) \; [support = 1\%, confidence = 50\%],$$

Multi-dimensional association rule?

$$age(X, \text{``20..29''}) \wedge income(X, \text{``40K..49K''}) \Rightarrow buys(X, \text{``laptop''})$$
$$[support = 2\%, confidence = 60\%].$$

"Minimum support threshold and Minimum confidence threshold"

7

3. Classification and Regression for Predictive Analysis

Classification is the process of finding a model that describes and distinguishes data classes or data concepts

The model are derived based on the analysis of a set of training data. This model is used for predicting the class label of unknown dataset.

Ex:	80	80	80	Excellent
	70	70	70	Very Good
	60	60	60	Good
	90	80	90	Excellent
	75	70	60	Very Good
	60	62	61	Good
	65	61	62	Unknown?
	82	80	85	Unknown?

3. Classification and Regression for Predictive Analysis (Contd..)

Forms of Classification
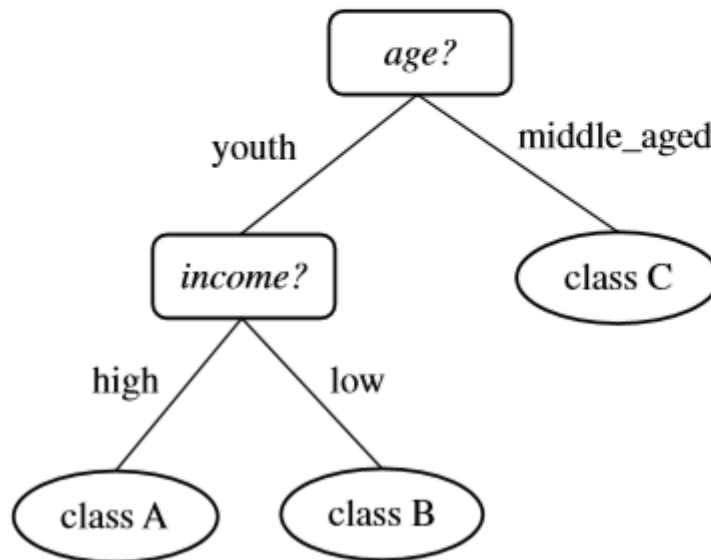
1. **IF-THEN Rules**
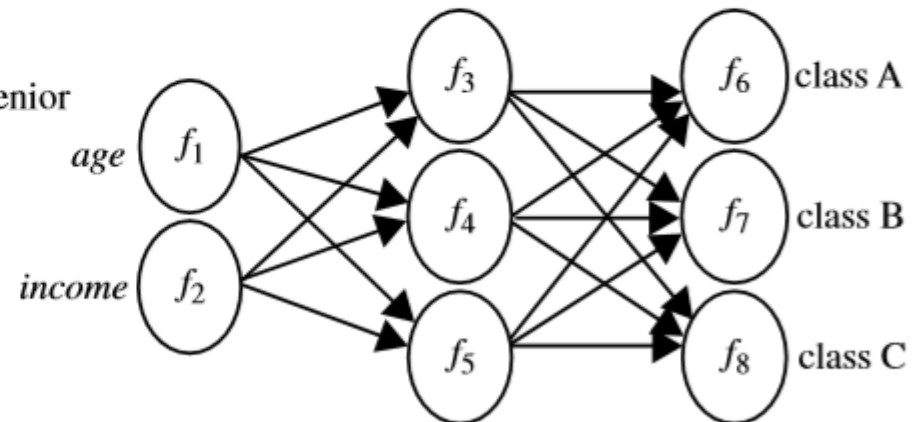2. **Decision Tree**
3. **A Neural Network**

3. Classification and Regression for Predictive Analysis (Contd..)

age(X, "youth") AND income(X, "high") ⟶ class(X, "A")

age(X, "youth") AND income(X, "low") ⟶ class(X, "B")

age(X, "middle_aged") ⟶ class(X, "C")

age(X, "senior") ⟶ class(X, "C")

(a)

(b)

(c)

1. What are the kinds of pattern analysis techniques or data mining functionalities
2. Define class/ concept descriptions?
3. Define data characterization?
4. Define data discrimination?
5. What is classification and how can you define the classification model?
6. What are the various forms of classification?

Thank You