

การทํานาย



CS 412 Intro. to Data Mining

Chapter 8. Classification: Basic Concepts

Jiawei Han, Computer Science, Univ. Illinois at Urbana-Champaign, 2017



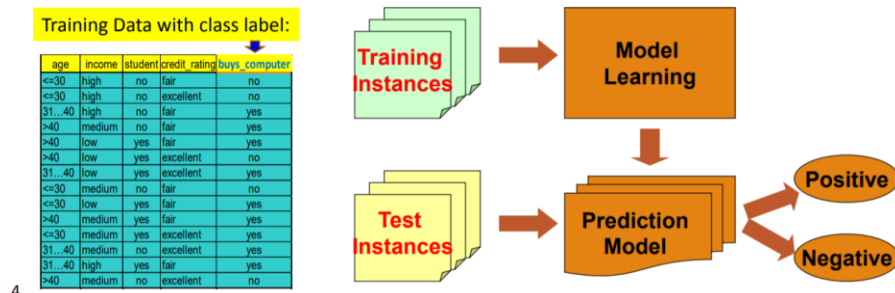
แบ่งได้เป็นสองกลุ่ม

Supervised learning (classification)

การสร้างโมเดลแบบมีผู้สอน

□ Supervised learning (classification)

- Supervision: The training data such as observations or measurements are accompanied by **labels** indicating the classes which they belong to
- New data is classified based on the models built from the training set



เก็บข้อมูลของลูกค้าที่ซื้อคอมพิวเตอร์ แต่ละคอลัมน์ ก็จะแบ่งเป็นอายุ , รายได้ , สถานะ,เครดิตดีแค่ไหน, ตัดสินใจซื้อ/ไม่ซื้อ เป็นโมเดลแบบมีผู้สอน

Unsupervised learning (clustering) คือโมเดลแบบไม่มีผู้สอน

มีการจัดกลุ่มว่าใครจะอยู่กับใครก็ได้ ไม่มีจุดมุ่งหมายตั้งแต่ตอนต้น

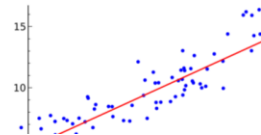
Prediction Problems: Classification vs. Numeric Prediction

Classification

- Predict categorical class labels (discrete or nominal)
- Construct a model based on the training set and the **class labels** (the values in a classifying attribute) and use it in classifying new data

Numeric prediction

- Model continuous-valued functions (i.e., predict unknown or missing values)
- Typical applications of classification
 - Credit/loan approval
 - Medical diagnosis: if a tumor is cancerous or benign



เราทำการสร้างโมเดลฟิเจอร์ก็เพื่อที่จะมาใช้ในการทำนายคำตอบ

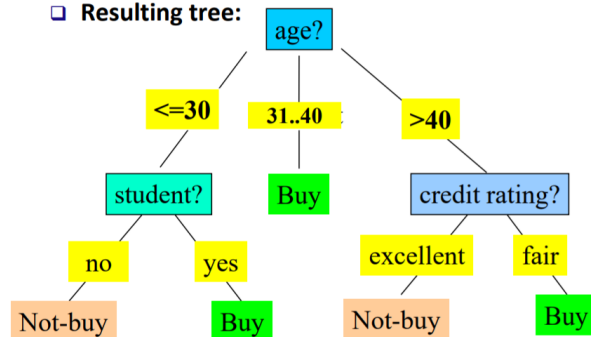
ต้นไม้ทำนายคำตอบ

Decision Tree Induction: An Example

Decision tree construction:

- A top-down, recursive, divide-and-conquer process

Resulting tree:



Training data set: Who buys computer?

age	income	student	credit rating	buys computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

Note: The data set is adapted from "Playing Tennis" example of R. Quinlan

X = Feature

Y = label

การทำแผนผังต้นไม้ จะเลือกตั้งแต่การแบ่งกลุ่มของช่วงอายุ โดยการสร้าง Decision tree จะเริ่มจากรากไปใบ ถ้าถูกจัดกลุ่มเรียบร้อยแล้วจะไม่ทำการแยกต่อ

