CHAPTER 1:

INTRODUCTION TO MACHINE LEARNING

Big Data

- Widespread use of personal computers and wireless communication leads to "big data"
- We are both producers and consumers of data
- Data is not random, it has structure, e.g., customer behavior
- We need "big theory" to extract that structure from data for

  (a) Understanding the process

  (b) Making predictions for the future

Why "Learn" ?

- To be intelligent, a system should have an ability to learn and to adapt to the changes of environment.
- A model of environment defined with some parameters → Need to optimize the parameters
- Machine Learning is programming computers to optimize a performance criterion using example data or past experience.
- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
    - E.g.) no need to "learn" to calculate payroll
  - Humans are unable to explain their expertise (speech recognition)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases (user biometrics)

What we talk about Learning?

- Learning general models from a data of particular examples.
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.
- Example in retail: Customer transactions to consumer behavior:

  People who bought "Blink" also bought "Outliers" (www.amazon.com)
- Build a model that is *a good and useful approximation* to the data.
- ML uses the theory of statistics in building mathematical model.

Data Mining

5

- Application of ML methods to large DB to construct a model.
- Retail: Market basket analysis, Customer relationship management (CRM)
- Finance: Credit scoring, fraud detection
- Manufacturing: Control, robotics, troubleshooting
- Medicine: Medical diagnosis
- Telecommunications: Spam filters, intrusion detection
- Bioinformatics: Motifs, alignment
- Web mining: Search engines
- ...

What is Machine Learning?

6

- Optimize a performance criterion using example data or past experience to learn the model.
- The model may be predictive or descriptive, or both.
- Role of Statistics: Inference from a sample
- Role of Computer science: Efficient algorithms to
  - Solve the optimization problem as well as store/process data.
  - Representing and evaluating the learned model for inference.

Applications

7

- Association
- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
- Reinforcement Learning

Learning Associations

8

- In finding association rule, learn a conditional probability, P(Y | X).
- Basket analysis:

  $P(Y \mid X)$ probability that somebody who buys X also buys Y where X and Y are products/services.

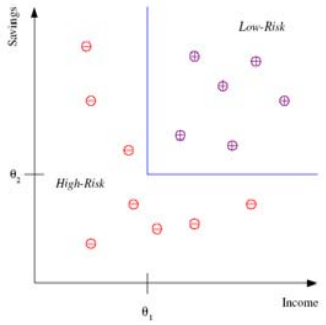
  Example: $P(\text{chips} \mid \text{beer}) = 0.7$

Classification

9

- $g: X^d \rightarrow Y$ – the output Y is one of a finite set of values.
- Example:

  Credit scoring
- Differentiating between low-risk and high-risk customers from their *income* and *savings*
- Discriminant: a function that separates the examples of different classes.
- Further prediction with discriminant.
- Classification ⇒ Association



Discriminant: IF *income* > θ_1 AND *savings* > θ_2
THEN low-risk ELSE high-risk

Classification: Applications

10

- Aka Pattern recognition
- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.

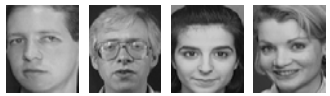
  acoustic signal → a word
- Medical diagnosis: From symptoms to illnesses
- Biometrics: Recognition/authentication of people using physical and/or behavioral characteristics: Face, iris, signature, etc.
- Knowledge extraction:
- Outlier/novelty detection:

Face Recognition

11

Training examples of a person

Test images



ORL dataset,
AT&T Laboratories, Cambridge UK

Regression

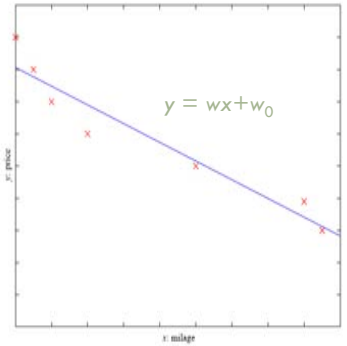
- g : X^d → Y where Y is a number.
- Example: Price of a usec car
- *x* : car attributes

  y : price

  $y = g(x \mid \theta)$

  g ( ) : model,

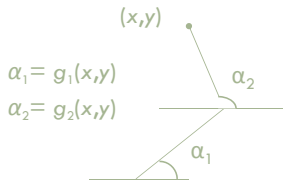
  θ : parameters

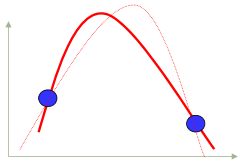


Regression Applications

13

- Navigating a car:
  - input: provided by sensors
  - output : Angle of the steering at each time.
- Kinematics of a robot arm





- Response surface design

Supervised Learning: Uses

14

- Prediction of future cases: Use the rule to predict the output for future inputs
- Knowledge extraction: The rule is easy to understand
- Compression: The rule is simpler than the data it explains
- Outlier detection: Exceptions that are not covered by the rule, e.g., fraud

Unsupervised Learning

15

- Learning *regularities* in the input.
- No output
- In statistics, density estimation.
- Clustering: Grouping similar instances
- Example applications
  - Customer segmentation in CRM (Customer Relationship Mgt.)
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs

Reinforcement Learning

16

- Learning a policy: A sequence of outputs
- No supervised output but delayed *reward*
- Credit assignment problem
- Game playing
- Robot in a maze
- Multiple agents, partial observability, ...

Resources: Journals & Data sets

17

- Journal of Machine Learning Research www.jmlr.org
- Machine Learning
- Neural Computation
- Neural Networks
- IEEE Trans on Neural Networks and Learning Systems
- IEEE Trans on Pattern Analysis and Machine Intelligence
- Journals on Statistics/Data Mining/Signal Processing/Natural Language Processing/Bioinformatics/…

Data

- UCI Repository: http://www.ics.uci.edu/~mlearn/MLRepository.html
- Statlib: http://lib.stat.cmu.edu/