

CHAPTER 1:

INTRODUCTION TO MACHINE LEARNING

國立雲林科技大學 資訊工程研究所

張傳育(Chuan-Yu Chang) 博士

Office: EB 212

TEL: 05-5342601 ext. 4516

E-mail: chuanyu@yuntech.edu.tw

Website: <http://MIPL.yuntech.edu.tw>

Big Data

2

- 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” ?

3

- 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),
 - Humans are unable to explain their expertise (face recognition)
 - Solution changes in time (routing on a computer network)
 - Solution needs to be adapted to particular cases (user biometrics)

What We Talk About When We Talk About “Learning”

4

- 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.

Data Mining

5

- **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.
- Role of Statistics: Inference from a sample
- Role of Computer science: Efficient algorithms to
 - Solve the optimization problem
 - Representing and evaluating the model for inference

Applications

7

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

Learning Associations

8

- Basket analysis:

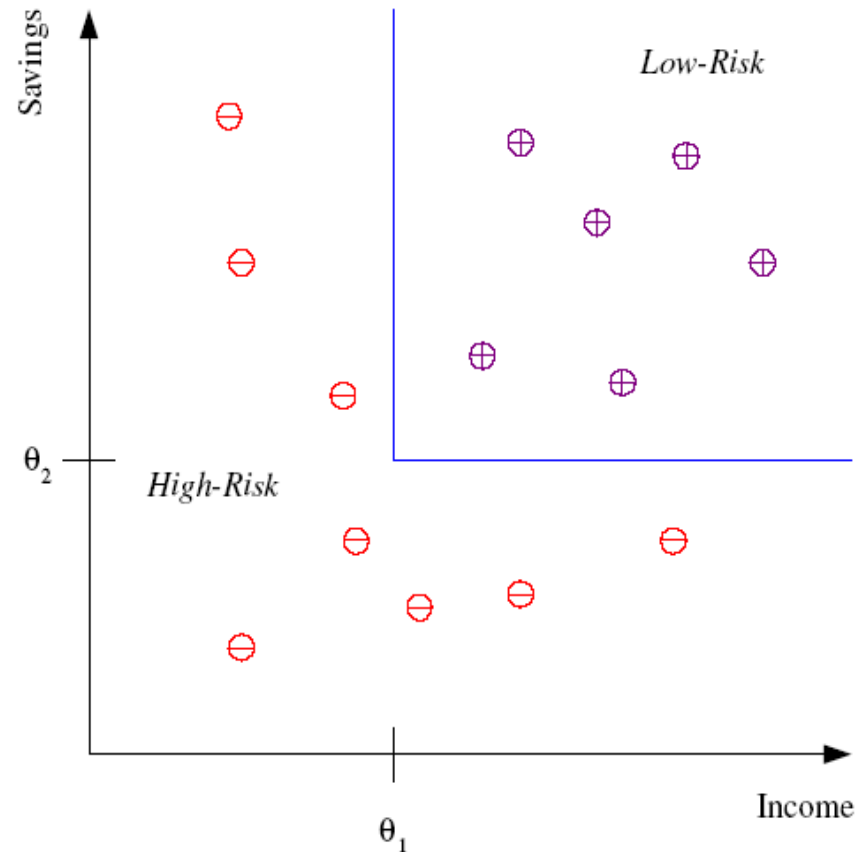
$P(Y | X)$ probability that somebody who buys X also buys Y where X and Y are products/ services.

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

Classification

9

- Example: Credit scoring
- Differentiating between **low-risk** and **high-risk** customers from their *income* and *savings*



Discriminant: IF $income > \theta_1$ AND $savings > \theta_2$
THEN **low-risk** ELSE **high-risk**

Classification: Applications

10

- Pattern recognition
- **Face recognition:** Pose, lighting, occlusion (glasses, beard), make-up, hair style
- **Character recognition:** Different handwriting styles.
- **Speech recognition:** Temporal dependency.
- **Medical diagnosis:** From symptoms to illnesses
- **Biometrics:** Recognition/authentication using physical and/or behavioral characteristics: Face, iris, signature, etc
- **Outlier/novelty detection:**

Face Recognition

11

Training examples of a person



Test images



ORL dataset,
AT&T Laboratories, Cambridge UK

Regression

□ Example: Price of a used car

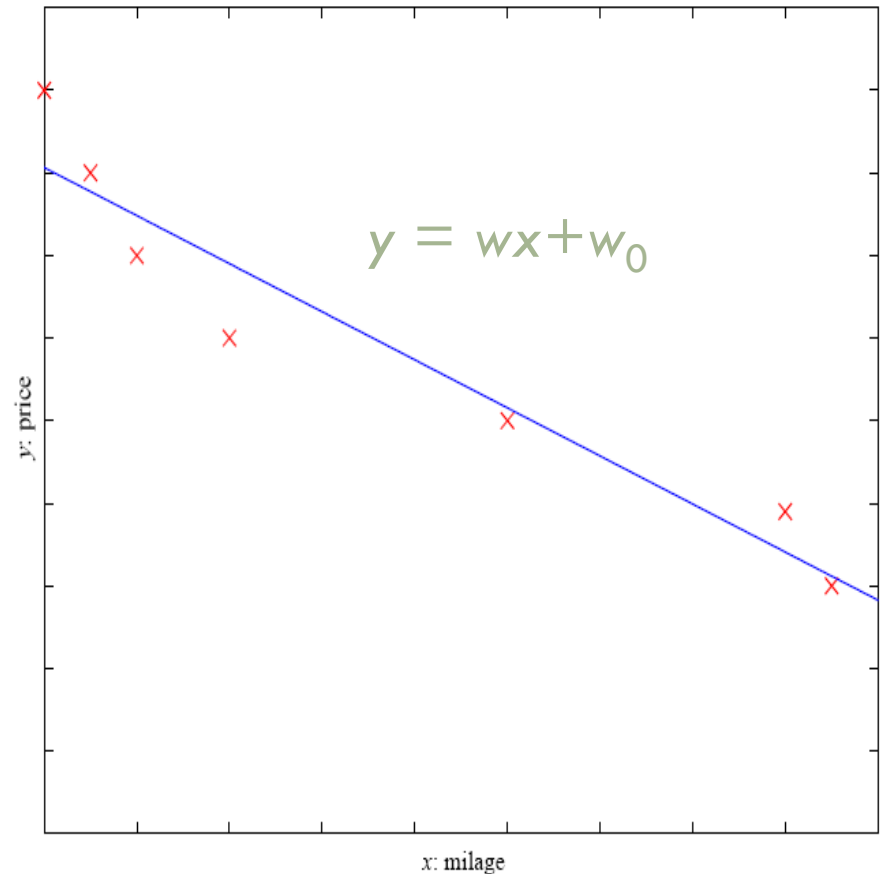
□ x : car attributes

y : price

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

$g(\cdot)$: model,

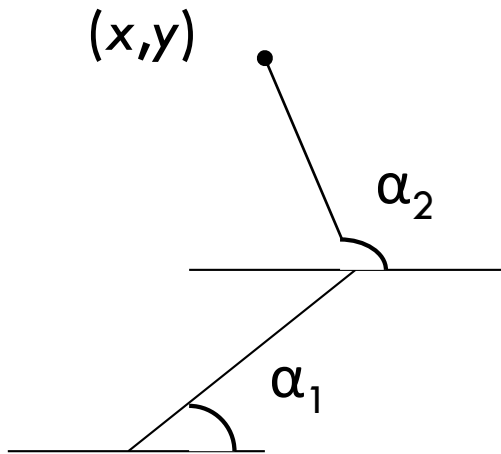
θ : parameters



Regression Applications

13

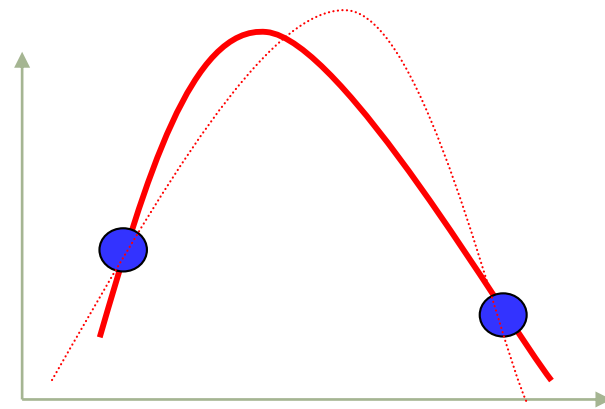
- Navigating a car: Angle of the steering
- Kinematics of a robot arm



$$\alpha_1 = g_1(x, y)$$

$$\alpha_2 = g_2(x, y)$$

- 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 “what normally happens”
- No supervisor
- To find the regularities in the input.
- Clustering: Grouping similar instances
- Example applications
 - Customer segmentation in CRM
 - 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: Datasets

17

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

Resources: Journals

18

- 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/...

Resources: Conferences

19

- International Conference on Machine Learning (ICML)
- European Conference on Machine Learning (ECML)
- Neural Information Processing Systems (NIPS)
- Uncertainty in Artificial Intelligence (UAI)
- Computational Learning Theory (COLT)
- International Conference on Artificial Neural Networks (ICANN)
- International Conference on AI & Statistics (AISTATS)
- International Conference on Pattern Recognition (ICPR)
- ...