SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY

COURSE NAME: 19TS622-MACHINE LEARNING

III YEAR /VI SEMESTER

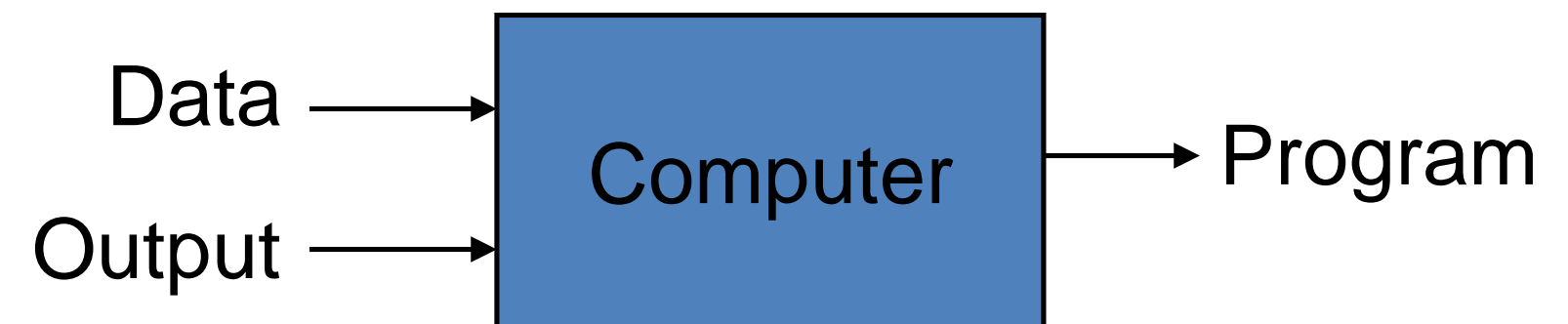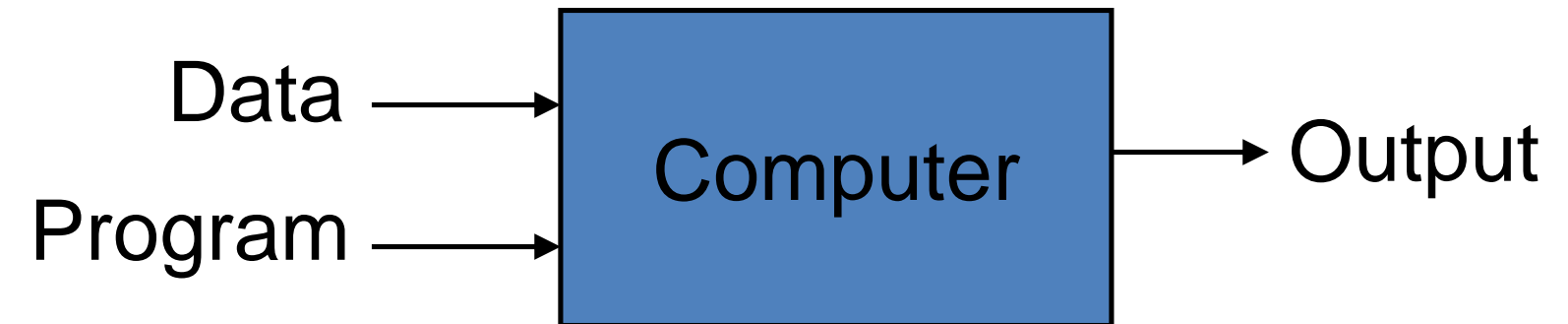Unit 1- INTRODUCTION

Topic 1: Machine Learning

Brain Storming

1. What is a Machine Learning?

2. What is all the excitement about?

Data ⟶ **Computer** ⟶ Output
Program ⟶

Data ⟶ **Computer** ⟶ Program
Output ⟶

Why "Learn" ?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.

- There is no need to "learn" to calculate payroll

- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
  - 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 When 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.

Data Mining

- 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?

- 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

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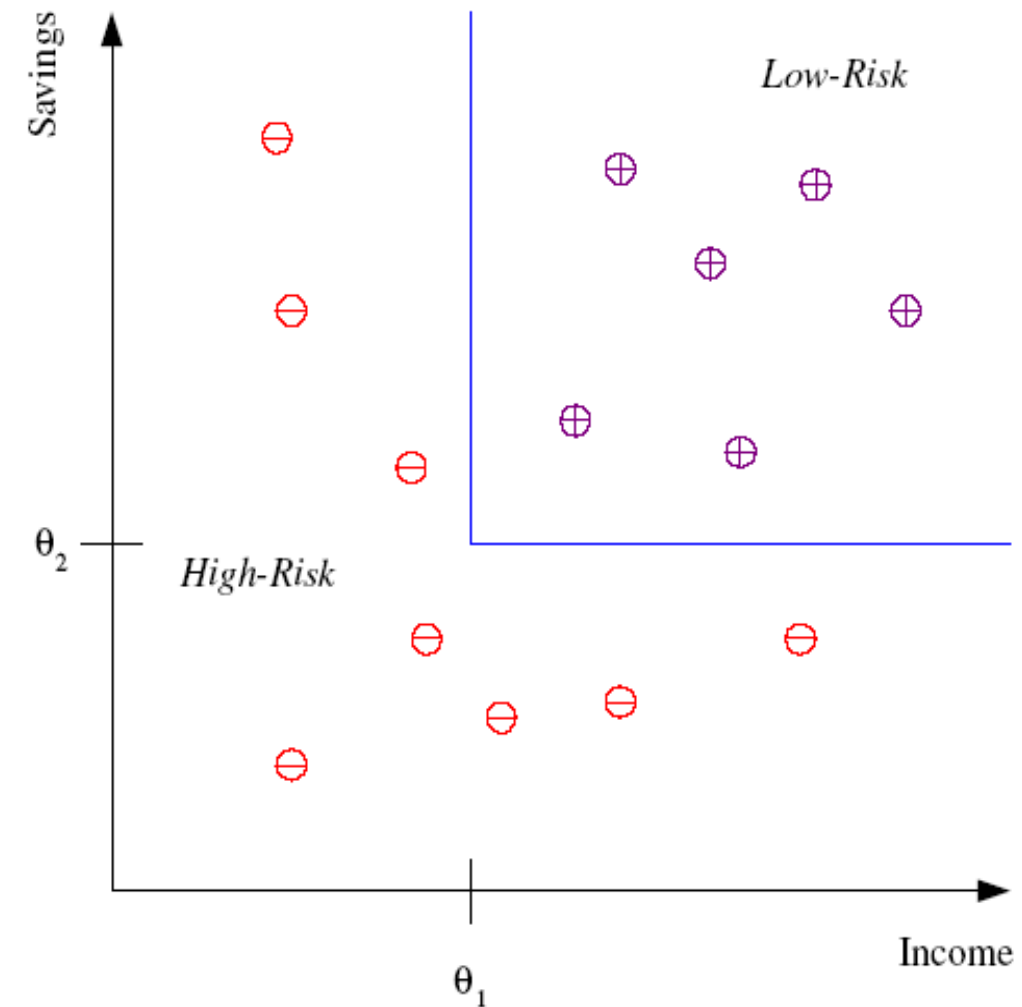- Unsupervised Learning
- Reinforcement Learning

Machine Learning/19TS622-Machine Learning / Dr.ARAVINDHAN K/CSE/SNSCE

- 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

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

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

Machine Learning/19TS622-Machine Learning / Dr.ARAVINDHAN K/CSE/SNSCE

Classification: Applications

- Aka 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
- ...

Face Recognition

Training examples of a person

Test images

ORL dataset,
AT&T Laboratories, Cambridge UK
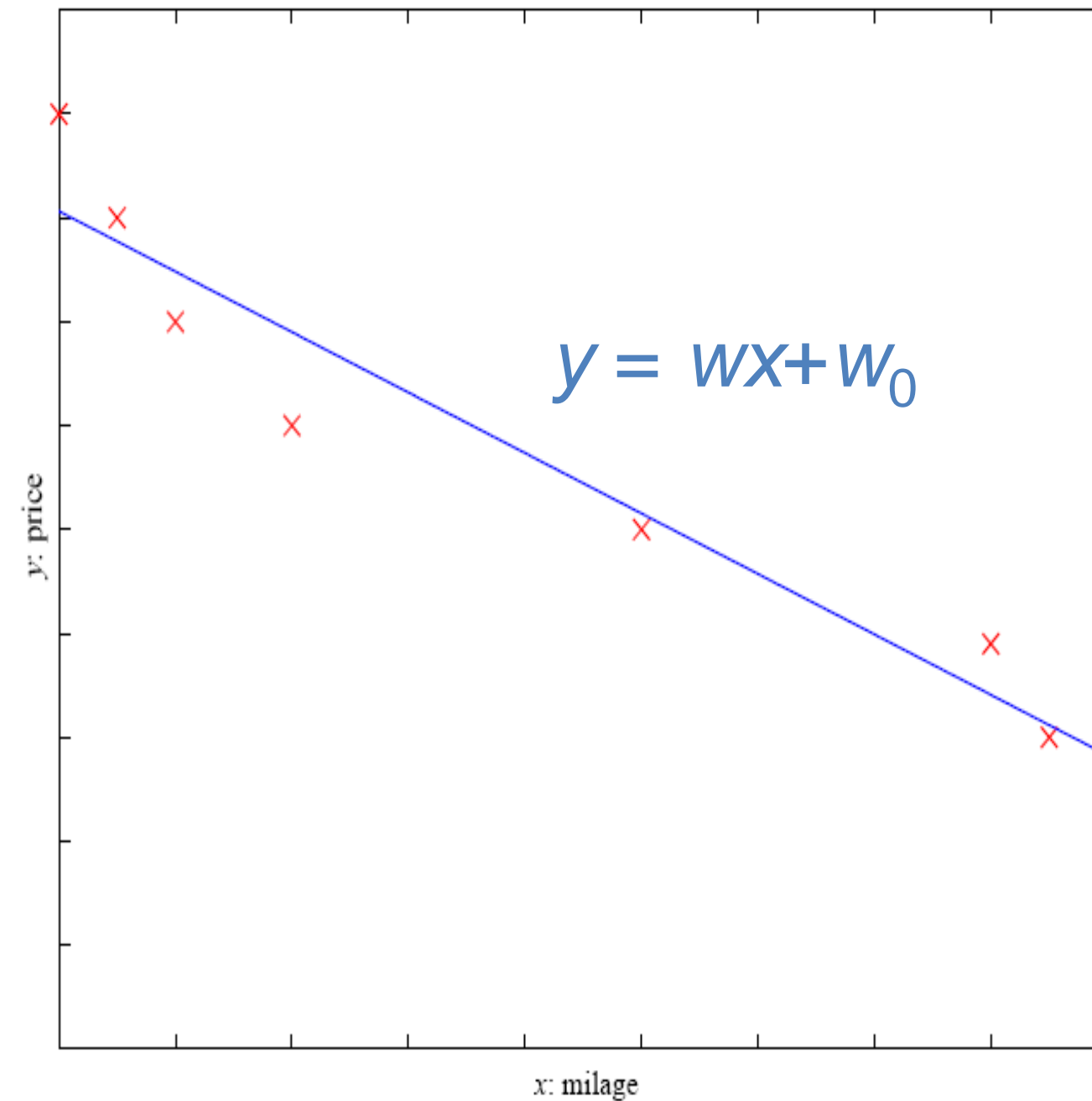
Regression

- Example: Price of a used car
- x : car attributes
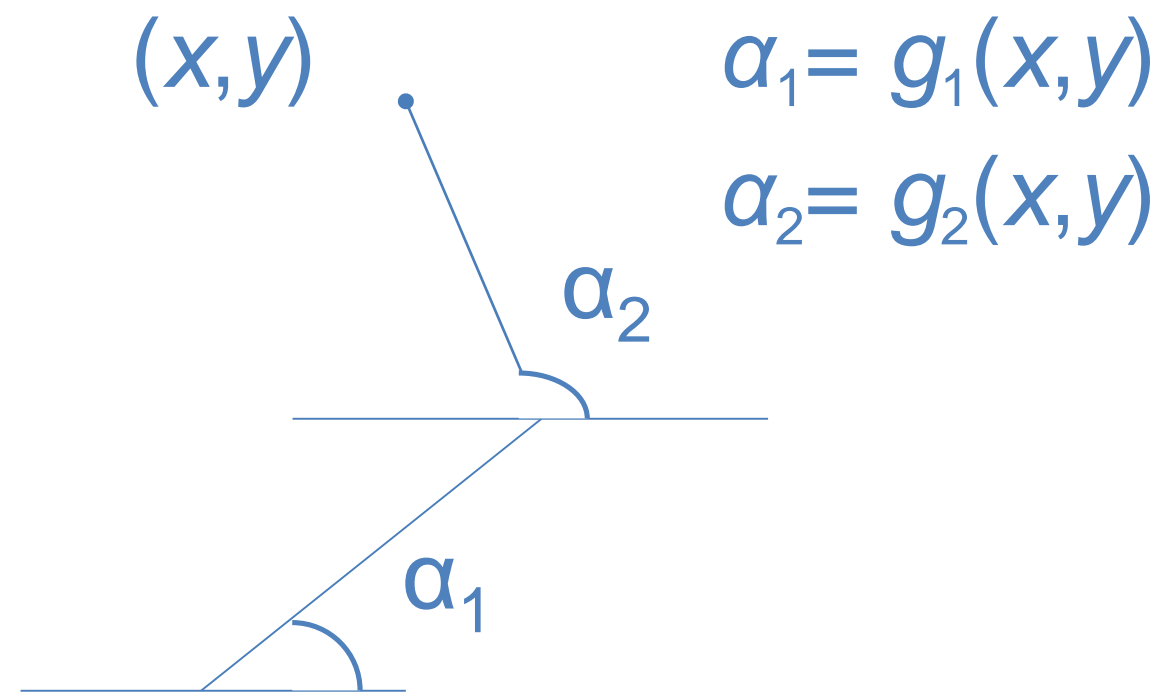
  y : price

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

  g ( ) model,
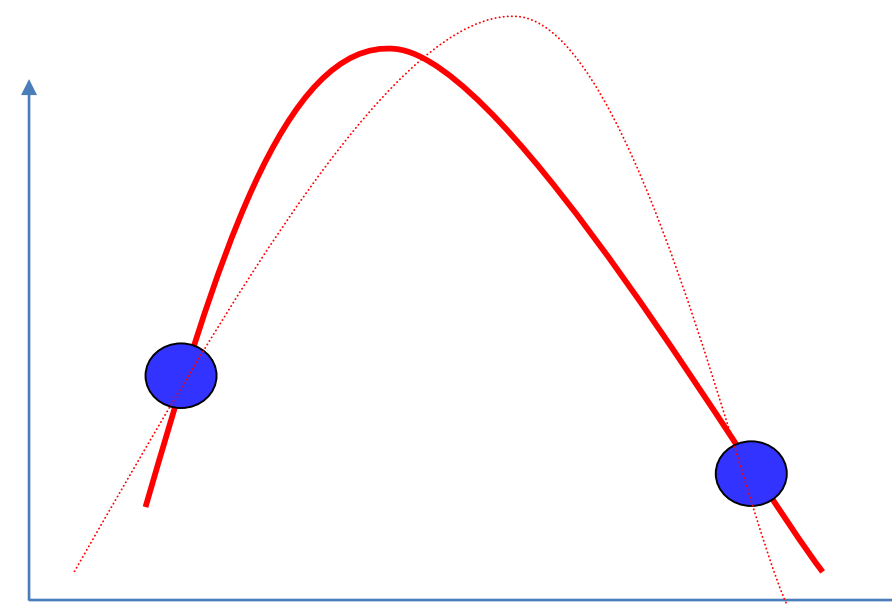
  θ parameters

$$y = wx + w_0$$

y: price

x: milage

Regression Applications

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

(x,y)

$\alpha_1 = g_1(x,y)$

$\alpha_2 = g_2(x,y)$

α_2

α_1

■ Response surface design

Machine Learning/19TS622-Machine Learning / Dr.ARAVINDHAN K/CSE/SNSCE

Supervised Learning: Uses

- **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

- Learning "what normally happens"
- No output
- Clustering: Grouping similar instances
- Example applications
  - Customer segmentation in CRM
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs

Reinforcement Learning

- 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

- UCI Repository: http://www.ics.uci.edu/~mlearn/MLRepository.html
- UCI KDD Archive: http://kdd.ics.uci.edu/summary.data.application.html
- Statlib: http://lib.stat.cmu.edu/
- Delve: http://www.cs.utoronto.ca/~delve/

Assessment 1

No, more like gardening

Seeds = Algorithms
Nutrients = Data
Gardener = You
Plants = Programs

References

TEXT BOOKS

1. Alpaydin Ethem, "Introduction to Machine Learning", MIT Press, Second Edition, 2010

REFERENCES

1. AurélienGéron, Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 2nd Edition, o'reilly, (2017)

2. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, "An Introduction to Statistical Learning: with Applications in R", Springer; First Edition 2013.

3. P. Flach, —Machine Learning: The art and science of algorithms that make sense of data,Cambridge University Press, 2012.

4. Tom M. Mitchell, "Machine Learning", McGraw-Hill Education (India) Private Limited, 2013.

Thank You