UMBC

AN HONORS UNIVERSITY IN MARYLAND

IS 709/809: Computational Methods in IS Research

Introduction to Machine Learning

Nirmalya Roy

Department of Information Systems

University of Maryland Baltimore County

Computational Methods (contd.)

Data Science/
Machine Learning/
Statistics

Machine Learning?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience
  - We don't have a specific algorithm to identify spam emails
- There is no need to "learn" to calculate payroll
  - No need to learn to sort the numbers
- 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
  - Consider a superstore example: we don't know exactly what people are going to buy
  - If we knew we can write algorithm & code
- Data is cheap and abundant (data warehouses, data marts)
  - knowledge is expensive and scarce
- Example: Customer transactions to consumer behavior:
  - *People who buy "pasta" may also buy "pasta sauce"*

What is learning?

- A process that explains the data we observe
  - Details of the process underlying the generation of data is unknown
  - Not completely random
- Build a model that is *a good and useful approximation* to the data
  - detect certain patterns or regularities
  - help us to understand process
  - use those patterns to make predictions

Data Mining

- Application of machine learning methods to large databases is called 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
  - It is not just a database problem, it is also a part of artificial intelligence
- A system with *ability to learn* in changing environment & adapt to such changes
  - Designer does not need to foresee and provide all possible solutions
- Theory of Statistics: Inference from a sample
- Role of Computer Science: Design efficient algorithms to
  - Solve the optimization problem
  - Represent and evaluate the model for inference

Machine Learning Applications

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

Learning Associations

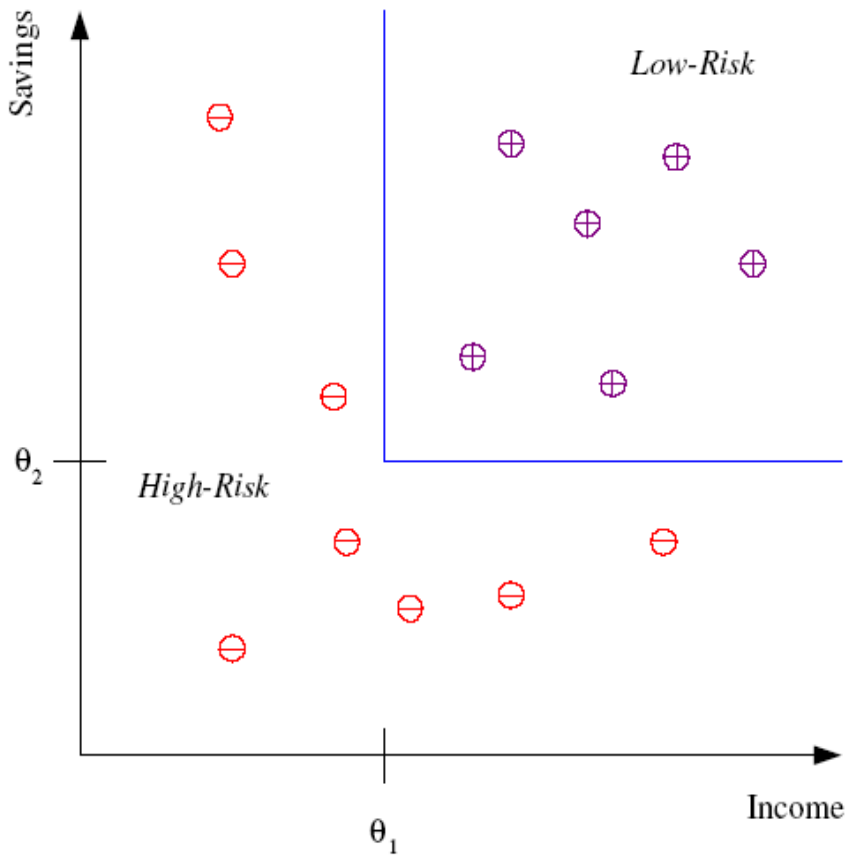
- Basket analysis:

  $P(Y \mid X)$ 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

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
- Knowledge extraction, compression, outlier detection….

Face Recognition

Training examples of a person

Test images

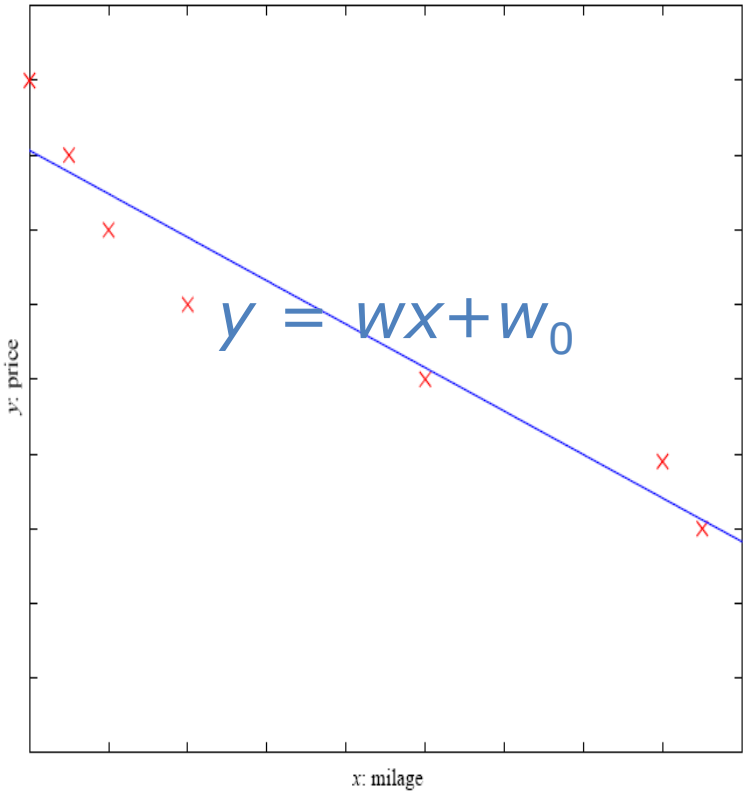
ORL dataset,
AT&T Laboratories, Cambridge UK

Regression

- Example: Price of a used car
- x : car attributes; y : price
- Collect training data
- ML program fits a function to this data to learn y
- *Assume a model*

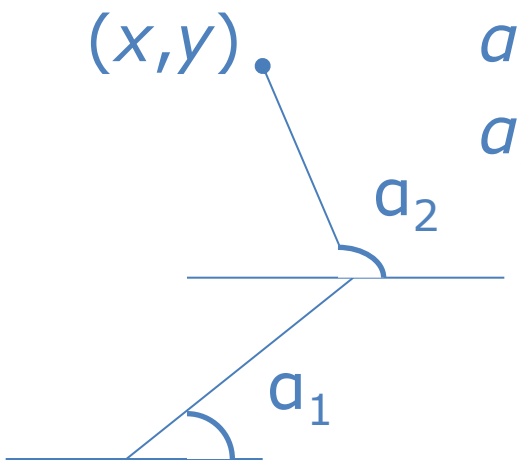
  $y = g(x \mid \theta)$

  $g(\ )$ model, θ parameters
- Regression and classification: Supervised problem?



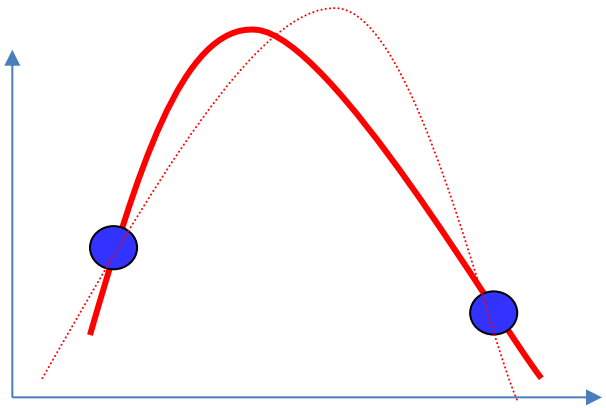
Regression Applications

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



$a_1 = g_1(x,y)$

$a_2 = g_2(x,y)$



- Response surface design

Supervised Learning

- 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

- Supervised learning:
  - Learn a mapping from the input to an output
  - Correct values of the output provided by a supervisor
- Unsupervised learning
  - No such supervisor
  - Only have input data, no output
  - Find the regularities in the input
  - Exploit the structure of the input space
  - Certain patterns occur more often than others
  - *Density estimation*
    - Example: clustering

Unsupervised Learning

- 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
  - A sequence of good actions to reach the goal
  - An action is good if it is part of a good policy
  - Assess the goodness of policies and learn from past
  - No supervised output but delayed reward
- Game playing
  - Single move by itself is not that important
  - Sequence of right moves that is good
- Robot navigating
  - Partial observability, ...May not know its exact location, but only that there is a wall to its left
  - Multiple agents (team of robots playing soccer)

ML 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/

ML Resources: Journals

- Journal of Machine Learning Research www.jmlr.org
- Machine Learning
- Neural Computation
- Neural Networks
- IEEE Transactions on Neural Networks
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Annals of Statistics
- Journal of the American Statistical Association
- ...

ML Resources: Conferences

- 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)
- ...

Questions

?