CHAPTER 1:

Introduction

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 "Da Vinci Code" also bought "The Five People You Meet in Heaven" (www.amazon.com)

- Build a model that is *a good and useful approximation* to the data.

Data Mining/KDD

- Definition := *"KDD is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data"* (Fayyad)

- **Applications:**
  - Retail: Market basket analysis, Customer relationship management (CRM)
  - Finance: Credit scoring, fraud detection
  - Manufacturing: Optimization, troubleshooting
  - Medicine: Medical diagnosis
  - Telecommunications: Quality of service optimization
  - Bioinformatics: Motifs, alignment
  - Web mining: Search engines
  - ...

What is Machine Learning?

- Machine Learning
  - Study of algorithms that
  - improve their performance
  - at some task
  - with experience
- 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

Growth of Machine Learning

- Machine learning is preferred approach to
  - Speech recognition, Natural language processing
  - Computer vision
  - Medical outcomes analysis
  - Robot control
  - Computational biology
- This trend is accelerating
  - Improved machine learning algorithms
  - Improved data capture, networking, faster computers
  - Software too complex to write by hand
  - New sensors / IO devices
  - Demand for self-customization to user, environment
  - It turns out to be difficult to extract knowledge from human experts → *failure of expert systems in the 1980's.*

Applications

- Association Analysis
- Supervised Learning
  - Classification
  - Regression/Prediction
- 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 (chips | coke) = 0.7

- *Example of Market-Basket transactions*

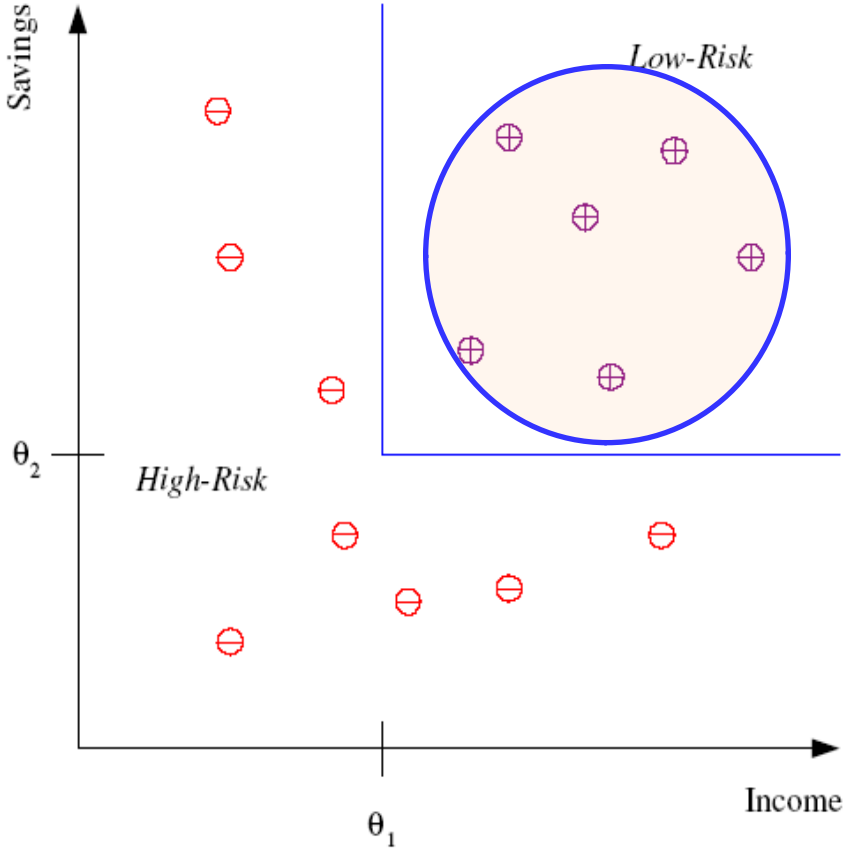
TID	*Items*
1	Bread, Milk
2	Bread, Diaper, milk, Eggs
3	Milk, Diaper, milk, Coke
4	Bread, Milk, Diaper, milk
5	Bread, Milk, Diaper, Coke

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
- Model

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.
  - Use of a dictionary or the syntax of the language.
  - Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- Medical diagnosis: From symptoms to illnesses
- Web Advertizing: Predict if a user clicks on an ad on the Internet.

Face Recognition

- Training examples of a person

- Test images

- AT&T Laboratories, Cambridge UK
- http://www.uk.research.att.com/facedatabase.html
- *

Prediction: Regression

Example: Price of a used car

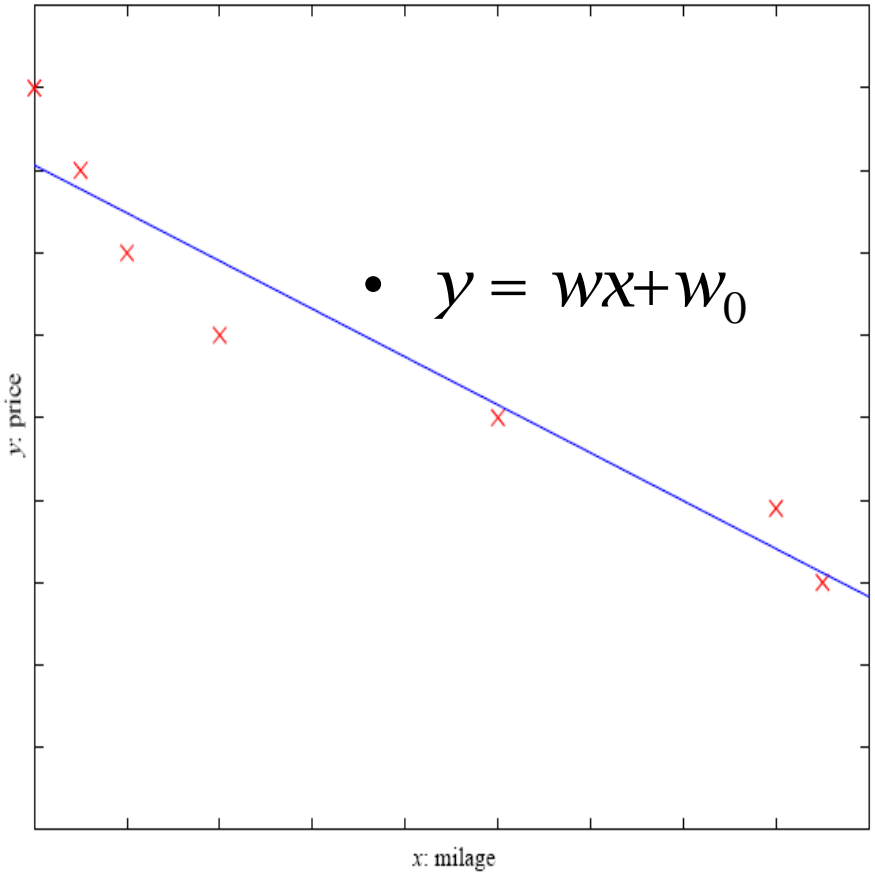
x : car attributes

y : price

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

$g(\ )$ model,

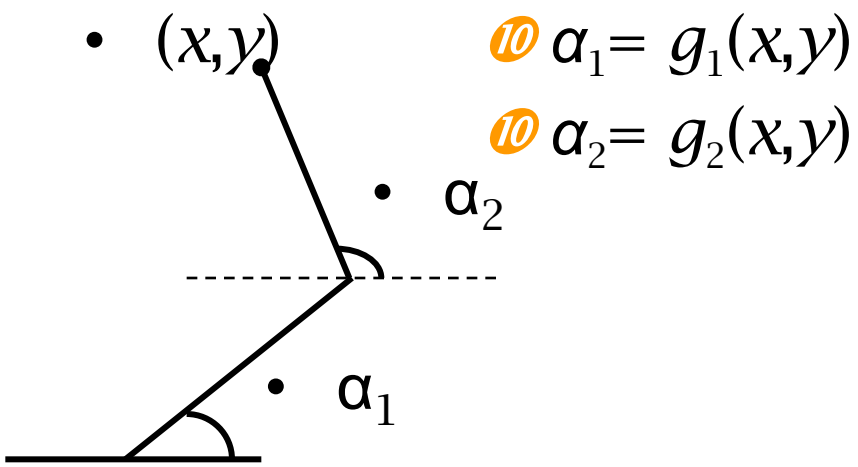
θ parameters



- Y: Price
- $y = wx+w_0$
- X : mileage
- *

RegressionApplications

- Navigating a car: Angle of the steering wheel (CMU NavLab)
- Kinematics of a robot arm



- $\alpha_1 = g_1(x,y)$
- $\alpha_2 = g_2(x,y)$

SupervisedLearning:Uses

- Example: decision trees tools that create rules
  - 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

UnsupervisedLearning

- Learning “what normally happens”
- No output
- Clustering: Grouping similar instances
- Other applications: Summarization, Association Analysis
- Example applications
  - Customer segmentation in CRM
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs

ReinforcementLearning

- Topics:
  - Policies: what actions should an agent take in a particular situation
  - Utility estimation: how good is a state (→used by policy)
- No supervised output but delayed reward
- Credit assignment problem (what was responsible for the outcome)
- Applications:
  - 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/

Resources: Journals

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

Resources:Conferences

- International Conference on Machine Learning (ICML)
- European Conference on Machine Learning (ECML)
- Neural Information Processing Systems (NIPS)
- Computational Learning
- International Joint Conference on Artificial Intelligence (IJCAI)
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)
- IEEE Int. Conf. on Data Mining (ICDM)

From book: Machine Learning, Tom Mitchel

What is the Learning Problem?

Learning = Improving with experience at some task

- Improve over task T,
- with respect to performance measure P,
- based on experience E.

E.g., Learn to play checkers

- T: Play checkers
- P: % of games won in world tournament
- E: opportunity to play against self