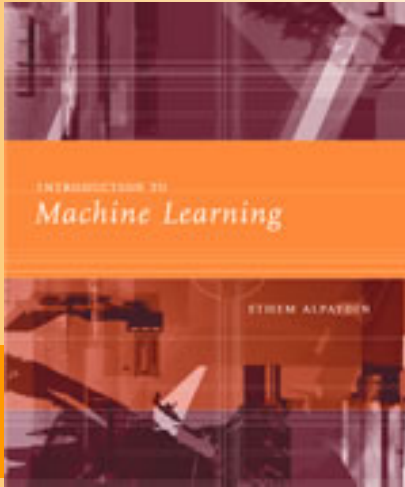


# *Lecture Slides for*



INTRODUCTION TO

# *Machine Learning*

ETHEM ALPAYDIN

© The MIT Press, 2004

*alpaydin@boun.edu.tr*

*<http://www.cmpe.boun.edu.tr/~ethem/i2ml>*



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*

- **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?*

- 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

# *Learning Associations*

- Basket analysis:

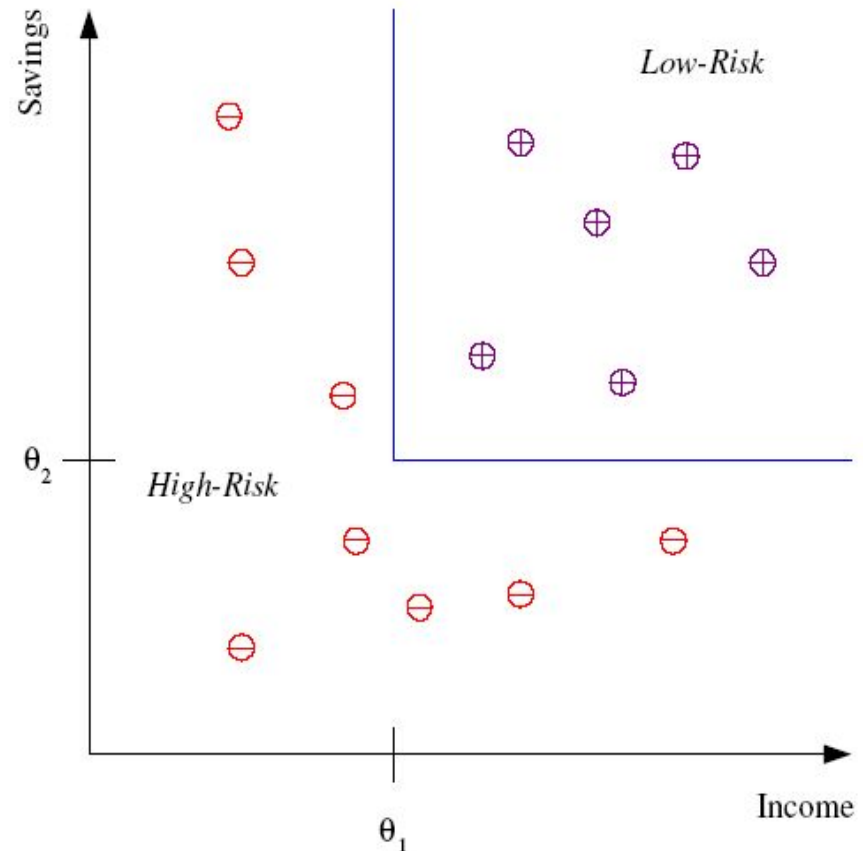
$P(Y | X)$  probability that somebody who buys  $X$  also buys  $Y$  where  $X$  and  $Y$  are products/services.

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



# Classification

- 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*

- 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
- ...

# Face Recognition

Training examples of a person



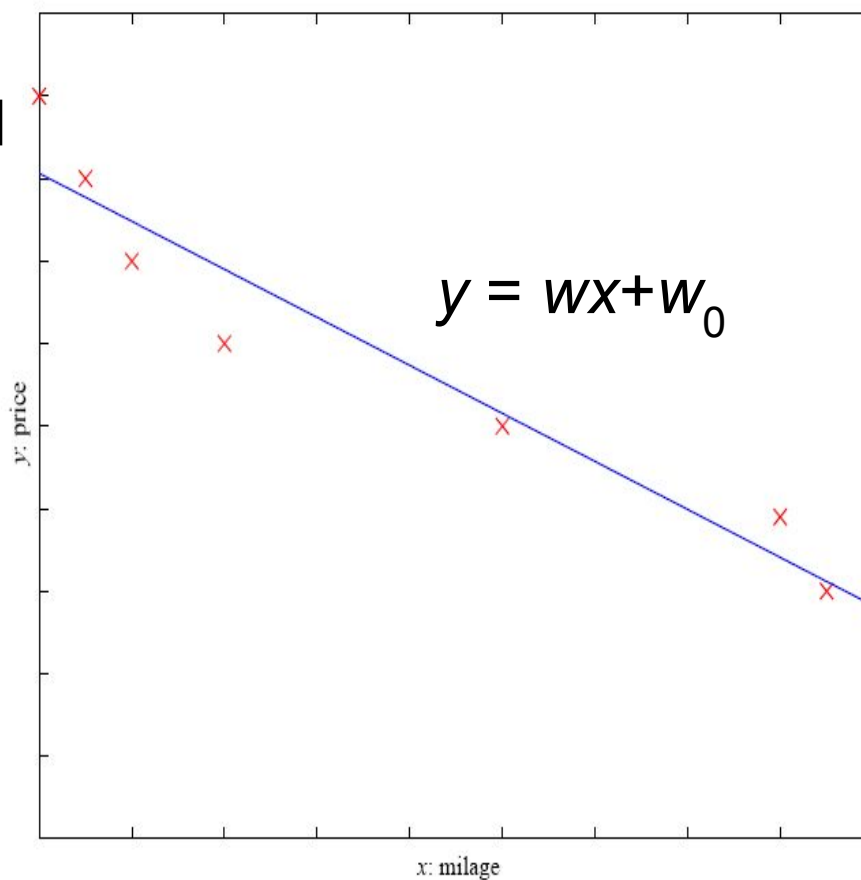
Test images



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

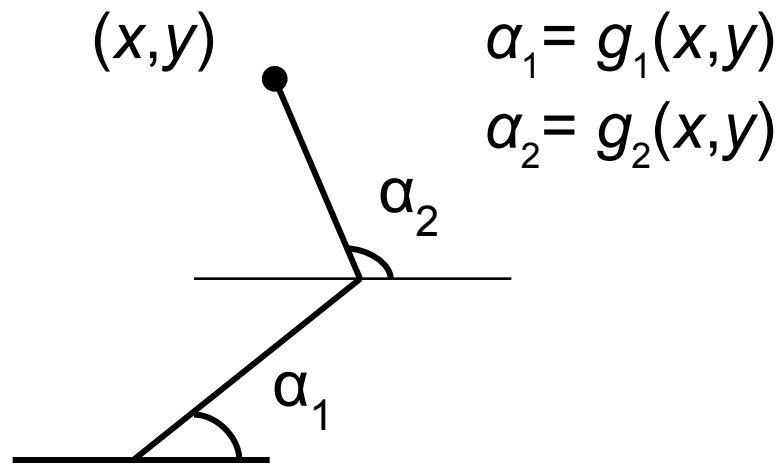
# Regression

- Example: Price of a used car
- $x$  : car attributes  
 $y$  : price  
 $y = g(x | \theta)$   
 $g(\cdot)$  model,  
 $\theta$  parameters

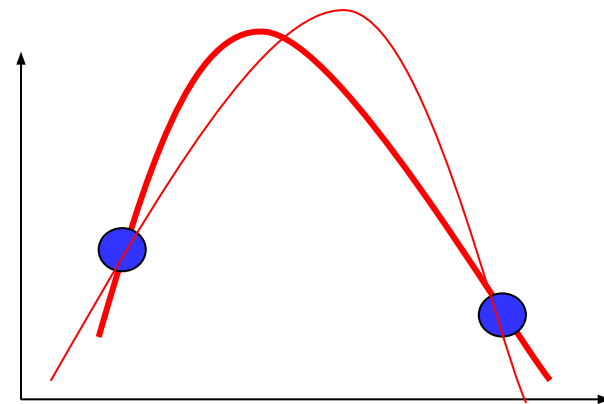


# Regression Applications

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



- Response surface design



# *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/>

# *Resources: Journals*

- Journal of Machine Learning Research [www.jmlr.org](http://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
- ...

# Resources: Conferences

- International Conference on Machine Learning (ICML)
  - ICML05: <http://icml.ais.fraunhofer.de/>
- European Conference on Machine Learning (ECML)
  - ECML05: <http://ecmlpkdd05.liacc.up.pt/>
- Neural Information Processing Systems (NIPS)
  - NIPS05: <http://nips.cc/>
- Uncertainty in Artificial Intelligence (UAI)
  - UAI05: <http://www.cs.toronto.edu/uai2005/>
- Computational Learning Theory (COLT)
  - COLT05: <http://learningtheory.org/colt2005/>
- International Joint Conference on Artificial Intelligence (IJCAI)
  - IJCAI05: <http://ijcai05.csd.abdn.ac.uk/>
- International Conference on Neural Networks (Europe)
  - ICANN05: <http://www.ibspan.waw.pl/ICANN-2005/>
- ...