Welcome to CMPS 242: Machine Learning

- Instructor: David Helmbold, dph@soe.ucsc.edu
- Web page: www.soe.ucsc.edu/classes/cmps242/Winter09/
- Text: Introduction to Machine Learning, Alpaydin

Administrivia

- Sign up sheet (enrollment)
- Evaluation:
  - Group Homework 25%
  - Late midterm 35%
  - Projects (group?) 40%
- Pictures
- Expectations/Style
  - Reading assignments
  - Attendance/participation
  - My hearing/writing
  - Academic honesty

Topics:
- Introduction (ch 1 and 2)
- Bayesian learning and parameter estimation (ch 3, 4, 5)
- Decision Trees (ch 9) and Art. Neural Networks (ch 11)
- Linear Discrimination (ch 10), the Perceptron algorithm (ch 11)
- Support vector machines (ch 10.9)
- Instance based learning (nearest neighbor) (ch 8)
- Boosting (AdaBoost) (ch 15)
- Clustering, EM Algorithm and K-means (ch 7)
- On-line prediction (Blum survey)
- Feature selection (ch 6)
- Concept learning and PAC model and generalization bounds
- Evaluating Hypotheses and model selection (ch 14) (if time)

INTRODUCTION TO Machine Learning

ETHEM ALPAYDIN
© The MIT Press, 2004
(modified by DPH, fall 2006)
alpaydin@boun.edu.tr
http://www.cmpc.boun.edu.tr/~ethem/i2ml

Why “Learn”? 

- Machine learning is programming computers to optimize a performance criterion using example data or past experience (inference in statistics)
- 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 or customized to particular cases or users (user biometrics)

What We Mean When We Talk About “Learning”

- Learning general models from a set of particular examples
- Data is often 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 “Angels and Demons” (www.amazon.com)
- Build a model that is a good and useful approximation to the data.
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

Stat. Machine learning is not:
- Cognitive science (how people think/learn)
- Teaching computers to think
But is related to:
- Statistics
- Data Mining - Knowledge Discovery in Databases
- Control theory
- part of AI, but not “traditional” AI

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
- ...

Supervised Batch Learning
- Assume distribution over a space of abstract things
- Get instance \( x \) by drawing things from distribution and recording measurements.
- Teacher labels instances making examples \((x, r)\)
  - Or \((x, y)\)
- Set of labeled examples is the sample or training set
- Create hypothesis (rule) from sample
- Hypothesis predicts on new instances, scored by loss function

Batch Learning Framework
- \( P(x, r) \) Test point
- Training set \( (x, r) \)
- Learning algorithm hypothesis
- Prediction \( \hat{y} \)
- Goal: find hypothesis with small loss
- Loss function \( L(\hat{y}, r) \)

Supervised Learning (cont.)
- Classification: labels are nominal (unordered set)
  - e.g. {ham, spam}, {democrat, Republican, indep.}
- Binary Classification
- Regression: labels are numeric values (e.g. price of used car)
- Sometimes labels are probabilities
Examples

<table>
<thead>
<tr>
<th>Thing</th>
<th>Observations</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written Digit</td>
<td>Pixel array</td>
<td>Which digit?</td>
</tr>
<tr>
<td>Email message</td>
<td>Words, Subject, sender</td>
<td>Ham or Spam?</td>
</tr>
<tr>
<td>Customer</td>
<td>Recent purchase</td>
<td>Interest level in new product</td>
</tr>
<tr>
<td>Used car</td>
<td>Year, make, mpg, options</td>
<td>Price or value</td>
</tr>
</tbody>
</table>

Face Recognition

Training examples of a person (with neg. examples?)

Test images

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

Regression

Example: Price of a used car

\[ y : \text{price} \]

\[ y = g(x | \theta) \]

\[ g(\cdot) \text{ model (e.g. linear)} \]

\[ \theta \text{ parameters (w, w_0)} \]

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 and data entry errors

Other kinds of supervised learning

- Reinforcement learning - learning a policy for influencing or reacting to environment
  - Policy maps "states" to "actions"
  - No supervised output, but delayed rewards
  - Credit assignment problem
  - Game playing/robot in a maze, etc.
- On-line learning: predict on each instance in turn
- Semi-supervised learning uses both labeled and unlabeled data

Unsupervised Learning

- No labels
- Learning "what normally happens"
- Clustering: Grouping similar instances
- Example applications
  - Segmentation in customer relationship mgmt
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs
  - Identifying unusual Airplane landings
Resources: Datasets

- Statlib: http://lib.stat.cmu.edu/

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

Resources: some 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 Joint Conference on Artificial Intelligence (IJCAI)
- International Conference on Neural Networks (ICANN)
- Knowledge Discovery and Data Mining (KDD)