Lecture 1: Introduction

Instructor: Yang Liu

General information

- Where: ECSN 2.311
- When: Mon and Wed 2:30-3:45pm
- Instructor:
  - Yang Liu, ECSS 3.402
    - Email: yangl@hl.tdallas.edu
    - Office hour: M: 3:45-4:30pm; W: 1:30-2:30pm
General information

- Course website: http://www.hlt.utdallas.edu/~yangl/cs6375
- TA: TBD
- Class notes and handouts: available from course web site
- Discussion and announcements: eLearning

Textbooks

- Reference textbooks:
  - Machine Learning, Tom Mitchell
  - Pattern Recognition and Machine Learning, Christopher Bishop
  - Introduction to Machine Learning, Ethem Alpaydin
  - Artificial Intelligence, Russell and Novig
  - Machine Learning: A Probabilistic Perspective, Kevin Murphy
- Other material available from the course webpage
Prerequisite

- CS 5345: Algorithm analysis and data structures
- Ability to program in C/C++, Java, or other languages
- Knowledge of math and probability/stats theory
- Ready to learn

Tentative grading policy

- Homework assignments: 35% (some programming, some exercises)
- Midterm (March 4, tentative): 30%
- Final exam (UTD schedule): 30%
- Closed book exam, one cheat sheet allowed

- Quiz and class participation: 5%
Course policies

- Homework policy
  - Collaboration is encouraged
  - But you have to write your own solutions/programs
- Late assignment policy
  - One day late: 85%
  - 2 days late: 70%
  - No assignment accepted after 2 days
- Re-grade policy
  - Requests must be made within one week of when the work was returned

What is “Machine Learning”

- “Field of study that gives computers the ability to learn without being explicitly programmed” [Arther Samuel]
- “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E” [Tom Mitchell]
What is “Machine Learning”

- Learn general models from a set of particular examples
  - Data is cheap and abundant (sometimes may not be the case); knowledge is expensive and scarce
    - E.g., use customer transactions to learn consumer behavior
  - Build a model that is a good and useful approximation of the data
- The ability to perform a task in a situation which has never been encountered before (learning = generalization)
- Different learning: classification, planning, finding patterns, etc.

Machine learning applications

- Speech recognition, natural language processing
- Computer vision
- Computational biology
- Association rules: somebody who buys X also buys Y
- Medical diagnosis: from symptoms to illness
- Security (detecting intrusion, worms, anomaly)
- More
Speech recognition

Recognize speech

Wreck a nice beach

Text classification

- Spam filtering

Dear Sir/Madam

I know that this letter may come to you as a surprise but due to the urgency of this transaction. First I must solicit your confidence in this transaction, this is by virtue of it’s nature as being utterly confidential and top secret…..

I am the manager of bill and exchange at the foreign remittance department of African Development bank (ADB). I came to know you in my private search for a reliable and reputable person to handle this Confidential Transaction, which involves the transfer of a huge sum of money to a foreign account requiring maximum confidence.

I am writing to you, following the impressive information received about you from the chambers of commerce. I believed that you are capable and reliable to champion this business opportunity. In my department we discovered an abandoned sum of $30m US dollars (Thirty million US dollars) …..
Text classification

- Sentiment analysis

Tell me a movie that is more famous than this. Tell me a movie that has had more parodies spun off its storyline than this. Tell me one movie that has been as quoted as a much as this. The answer is you can't. No movie has had as much of an impact as The Godfather has had ever since it was released.

The acting was simply amazing, what else could you say

Ambiguity resolution in language

- Word selection (speech recognition)
  - Can I have a peace of cake? Piece

- Word sense disambiguation
  - …Nissan car and truck plant is ..
  - Divide life into plant and animal kingdom

- Pronoun resolution
  - The dog bit the kid. He was taken to a vet.
IBM Watson Jeopardy

“Using machine learning, statistical analysis and natural language processing to find and understand the clues in the questions, Watson compared possible answers, by ranking its confidence in their accuracy, and responded – all in about three seconds.”
Fingerprint recognition

Image recognition

Note: using deep learning
Credit scoring

- Low-risk and high-risk customers based on income and savings

Play chess

May 11th, 1997

Computer won world champion of chess
(Deep Blue) (Garry Kasparov)
Unmanned car

Why study learning

- Important IT skills that employers look for (6-figure salary 😊)
- Lots of applications
- Computer systems with new capabilities
  - Develop systems that are too difficult or impossible to construct manually
  - Develop systems that can automatically adapt and customize themselves to the needs of the individual users through experience
  - Discover knowledge and patterns in databases, data mining
Why study learning

- Understand human and biological learning
- Time is right
  - Initial algorithms and theory in place
  - Growing amounts of data
  - Computational power available
  - Budding industry

Work in Machine Learning

- Artificial Intelligence
- Makes use of
  - Probability and statistics, linear algebra, calculus, optimization
- Related to
  - Philosophy, psychology, neurobiology, linguistics
- Has applications in
  - AI (natural language, vision, planning, HCI)
  - Computer science (compilers, systems, databases, software engineering, security)
Course overview

- Introduction to machine learning
- Supervised learning models and methods: decision trees, neural networks, nearest-neighbor algorithms, Bayesian learning, hidden Markov models (HMM), bagging, boosting, support vector machines (SVM)
- Unsupervised learning: clustering
- Reinforcement learning: Markov decision processes, Q-learning
- General techniques: feature selection, cross-validation, maximum likelihood estimation, gradient descent, expectation-maximization (EM)

Supervised learning

- Learning algorithms are given the correct target output
- Classification: learning to predict a discrete value from a predefined set of values
  - Whether people will like a movie or not, stock market goes up or down, an email is a spam
- Regression: learning to predict a continuous/real value
  - Stock price, driving
Example: dog or cat

- How does a child learn to distinguish a dog and a cat?
- Provide ‘training sets’, individually labeled ‘dog’ vs. ‘cat’, or other categories
- Learn the general rules that separate the ‘dog’ from ‘cat’
- Apply the learned rules to new situations

“Training” in machine learning

Training:
- Training examples (or instances)
- Training set
  
(picture of animal1, dog)
(picture of animal2, cat)
(picture of animal3, cat)

represented as a set of features/attributes

Learner

(Our Goal)
Function
(a.k.a. hypothesis concept description)

Special names:
classifier (if discrete)
regressor (if continuous)

Input: examples
Output: class

Learning algorithm

human
machine

label/class
(provided by teacher)
Evaluation in machine learning

Another example: work or play

<table>
<thead>
<tr>
<th>Outlook</th>
<th>temp</th>
<th>humidity</th>
<th>windy</th>
<th>Saturday</th>
<th>plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>hot</td>
<td>high</td>
<td>no</td>
<td>study</td>
<td></td>
</tr>
<tr>
<td>Sunny</td>
<td>hot</td>
<td>high</td>
<td>yes</td>
<td>study</td>
<td></td>
</tr>
<tr>
<td>Overcast</td>
<td>hot</td>
<td>high</td>
<td>no</td>
<td>play</td>
<td></td>
</tr>
<tr>
<td>Rain</td>
<td>mild</td>
<td>high</td>
<td>no</td>
<td>play</td>
<td></td>
</tr>
<tr>
<td>Rain</td>
<td>cool</td>
<td>normal</td>
<td>no</td>
<td>play</td>
<td></td>
</tr>
<tr>
<td>Rain</td>
<td>cool</td>
<td>normal</td>
<td>yes</td>
<td>study</td>
<td></td>
</tr>
</tbody>
</table>

- Goal: learn the target concept or function, f: day -> {work, play}
Comments on data

- Instances in the data set: each described by attributes or features (outlook, temp, humidity, windy in the Saturday plan example) and the class label (plan: work or study)

Evaluation issues

- Should test set be the same as training set?
  - Rote learning: memorization
  - Inductive learning: generalize from training examples

- Learning curves
  - Show how performance on test data varies with amount of training data
Inductive learning

- System tries to induce a “general rule” from a set of observed instances
- Given: collection of examples
- Return: a function $h$ (hypothesis) that approximates $f$ (target concept)

Inductive learning

- Inductive hypothesis: any hypothesis found to approximate the target function well over a sufficiently large set of training examples will also approximate the target function well over any other unobserved examples
- Assumptions for inductive learning algorithms:
  - Training sample represents the population
  - Input features permit discrimination
Learning ~ Search

- Learning is a search problem.
- Search through a space of possible hypotheses to find the one that best fits the available training data and other constraints.
- We will discuss various classifiers — search different hypothesis space.
- Need algorithms that can efficiently search large or infinite hypothesis space.

Unsupervised learning

- No training data.
- Input is a data set (say $\mathbb{R}^{D \times N}$).
- Output is a different representation of the input.
- Example: document clustering, dimensionality reduction, compression.
Resources

- Journals and conferences
  - JMLR, ML, IEEE transactions, ICML, NIPS, ...
- Data sets: UCI repository, UCI KDD, CMU Statlib, ...
- Tools
- Online lectures:
  - https://www.coursera.org/course/ml
  - https://www.coursera.org/course/machlearning