Theory and Applications of Machine Learning

University of Tartu
Spring 2008
Course Outline

• Lectures and discussions

• Course Readings

• Homeworks: Some questions and problems

• Reading presentation

• 2 Exams: One mid-term, one final

• Project
A few more course details

**Course Grade:** A (>90), B (>80), C (>70), D (>60), F (<60)

**Two equally weighted exams**
Exam 1 covers material from February to mid-April
Exam 2 covers material from mid-April to May

**Homeworks**
These will be assigned almost weekly and may contain bonus questions.

**Course Readings**
These are indicated in the detailed course description. They are an essential component of the course. You will also be required prepare a presentation based on a reading during the course for which peer grades will also contribute.

**Project:** More details in the course description
Some Course Topics

February - mid April

Introduction to machine learning terms
Optimization theory
Kernel Methods
Support vector machines (classification, regression, ranking)
Nu SVMs
Evaluative methods
Boosting
PCA and Data visualization
Clustering

Key dates:  March 20 - First Project Proposal due
            April 8 - Exam 1

**No class on March 6 and April 10 (I will be away)**
Second half of the Course

Mid-April to end of May

Manifold learning
KNN and novelty detection
Active learning
Probabilistic methods in Machine Learning
Data fusion
Multi-task learning
ML applications (biological, text mining, video scenes)

Key dates:
April 30 - Second Project Proposal due
May 27-29 - Project Presentations
June 3 - Exam 2
What is learning?

Learning is the process by which we acquire knowledge and information over time.

How does this happen?
Learning about the human learning machine

Cognitive Science studies the human brain … a mysterious mass containing an active network of over 100 billion neurons.
Learning about learning

The core principles of brain-based learning state that:

- The brain is a parallel processor, meaning it can perform several activities at once, like tasting and smelling.
- Learning engages the whole physiology.
- The search for meaning is innate.
- The search for meaning comes through patterning.
- Emotions are critical to patterning.

http://www.funderstanding.com/brain_based_learning.cfm

P. Agius – L1, Spring 2008
Patterns everywhere
Why patterns?

Predict the future …

Organization …
Pattern recognition

“... is the scientific discipline whose goal is the classification of objects into a number of categories or classes.”
[Pattern Recognition, Theodoridis & Koutroumbas, 1999]

“... concerned with identification of objects of known classes, or grouping of objects.”
[The Oxford Dictionary of Statistical Terms, Yadolah Dodge (Editor), 2003]

“Pattern recognition aims to classify data based on either a priori knowledge or on statistical information extracted from the patterns.”
[Wikipedia]
Some applications …

Understanding stock market behavior
Some applications …

Scheduling
Some applications …

Client categorization

Insurance agencies

Banks
Some applications …

Drug discovery
Some applications …

Understanding biological behavior at the genetic level
Some applications ...

Identifying diseases/cancer from various types of data
Some applications …

Google

Netflix

Ranking
Why \textbf{machine} learning?

Dietterich’s paper ‘Machine Learning’:

Masses of data for which no human expert exists. Machines can handle vast amounts of data.

Explaining expertise (Eg. How do we learn languages?)

Fluid data - managing rapid changes (Eg. Finance).

Customization (Eg. Mail filters) … just tune your parameters
Evaluating your model

The Bible Code is the title of a book by Michael Drosnin in which he claims that there is a code embedded in the Bible by God. The code is revealed by searching for equidistant letter sequences (ELS). The code is called the Bible Code or the Torah Code.
Some bible code ‘predictions’

Past predictions: Drosnin found the name “Edison” encoded near “light bulb” and “electricity”, and “Newton” encoded near “gravity”. The “prediction” of the 1995 assassination of Yitzhak Rabin. In 1994 Drosnin found Rabin’s name in the code, crossed by the words “assassin will assassinate”.

Coming attractions: An atomic holocaust ... a suicide bomber/terrorist from Bin Laden ignites an atomic attack on Jerusalem ... Barak, Sharon, Bin Laden and President G. W. Bush are all named ... Bible says they will be saying "peace" and "safety" and then sudden destruction will come upon them. Damascus will become a ruinous heap. There will be a great earthquake - "the big one" - in 2010. An asteroid or comet will be coming at us in 2012. World War III will start with a nuclear attack on Israel.

Some notation

In binary classification problems, the problem is to assign the data to one of two classes.

\[\{x_1, x_2, \ldots x_n\} \in X\]

\[\{y_1, y_2, \ldots y_n\} \in \{\pm 1\}\]
Supervised learning

Training data (aka sample)

Test data

Classification rate
% of test examples correctly classified 10/13

Misclassification rate
% of test examples incorrectly classified 3/13

Hyperplane
A few more terms

To find the hyperplane, we must define a similarity measure on X that is symmetric. We will be referring to this function as the kernel (more on this later).

Finding the hyperplane that minimizes the training error does not guarantee the function will perform well on test data.

In machine learning, the goal is to find a classifier (hyperplane) that generalizes well on future data, that is, we wish to minimize the misclassification error. Correctness of classification can be measured using a loss function.
Training and testing

Data -> Training data -> Building the Machine

Training data -> Test data -> Learning function

Evaluation

Repeat process multiple times, experiment with different test set sizes
Unsupervised learning

Unsupervised learning is a method of machine learning whereby the algorithm is presented with examples from the input space only and a model is fit to these observations. For example, a clustering algorithm would be a form of unsupervised learning. [Sewell, 2006]

"Unsupervised learning is a method of machine learning where a model is fit to observations. It is distinguished from supervised learning by the fact that there is no a priori output. In unsupervised learning, a data set of input objects is gathered. Unsupervised learning then typically treats input objects as a set of random variables. A joint density model is then built for the data set." [Wikipedia 2006]
In the **unsupervised learning problem**, we observe only the features and have no measurements of the outcome. Our task is rather to describe how the data are organized or clustered. [Hastie, Tibshirani and Friedman (2001), page 2]

"In unsupervised learning or clustering there is no explicit teacher ... different clustering algorithms lead to different clusters. Often the user will set the hypothesized number of different clusters ahead of time, but how should this be done? How do we avoid inappropriate representations?" [Duda, Hart and Stork (2001), page 17]

**Other unsupervised problems:**

- Projecting high-dimensional data into lower dimensional space
  (Principal Component Analysis [PCA],
  Multidimensional Scaling [MDS],
  Canonical Correlational Analysis [CCA])

- Novelty detection
The Curse of Dimensionality

High dimensionality in your input space can be a problem. Too much information! (too many suggested readings for this course!)

Example: Microarrays ($x_1 = \text{sample described by} > 22000 \text{ genes}$)

From “The curse of dimensionality”, Mario Koppen
Tom Mitchell’s white paper ‘The Discipline of Machine Learning’


**Computer Science**: “How can we build machines that solve problems, and which problems are inherently tractable/intractable?”

**Statistics**: “What can be inferred from data plus a set of modeling assumptions, with what reliability?”

**Machine Learning**: “How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?”
Applications of ML

Speech Recognition
Computer Vision
Bio-surveillance
Robot Control
Accelerating Empirical Sciences
Current research questions

Can unlabeled data be helpful for supervised learning?

How can we transfer what is learned for one task to improve learning in other related tasks?

What is the relationship between different learning algorithms, and which should be used when?

For learners that actively collect their own training data, what is the best strategy?

To what degree can we have both data privacy and the benefits of data mining?
 Longer term research questions

Can we build never-ending learners?

Can machine learning theories and algorithms help explain human learning?

Can we design programming languages containing machine learning primitives?

Will computer perception merge with machine learning?

Other Ethical questions
More ML applications

- Electric power load forecasting using KNN
- Automatic ‘help desk’ assistant
- Text mining
- Planning and scheduling
- Medical diagnosis

How is ML going to be useful to you?