T-61.5140 Machine Learning: Advanced Probablistic Methods

Jaakko Hollmén

Department of Information and Computer Science
Helsinki University of Technology, Finland
e-mail: `Jaakko.Hollmen@tkk.fi`
Web: `http://www.cis.hut.fi/Opinnot/T-61.5140/`

January 17, 2008

Course Organization: Personnel

Lecturer: Jaakko Hollmén, D.Sc.(Tech.)

- Lectures on Thursdays, from 10.15 - 12.00 in T3

Course Assistant: Tapani Raiko, D.Sc.(Tech.)

- Problem sessions on Fridays, from 10.15-12.00 in T3

For the schedule, holidays and special program, see

- `http://www.cis.hut.fi/Opinnot/T-61.5140/`

Course Material

Lecture slides and lectures

- Lecture notes (aid the presentation on the lectures)
- Lecture notes (contain extra material)

Course book

- Christopher M. Bishop: Pattern Recognition and Machine Learning, Springer, 2006
- Chapters 8,9,10,11, and 13 covered during the course

Problem sessions

- Problems and solutions
- Demonstrations

Participating on the Course

- Interest in machine learning
- Student number at TKK needed
- Course registration on the WebTopi System: `https://webtopi.tkk.fi`
- Prerequisites: T-61.3050 Machine Learning: Basic principles taught in Autumn by Kai Puolamäki and the necessary prerequisites for that course

Passing the Course (5 ECTS credit points)

- Attend the lectures and the exercise sessions for best learning experience :-)
- Browse the material before attending the lectures and complete the exercises
- Complete the term project requiring solving of a machine learning problem by programming
- Pass the examination, next exam scheduled: Thursday, 15th of May, morning
- Requirements: passed exam *and* a acceptable term project, bonus for active participation and excellent term project (+1)

Relation to Other Courses

This course replaces the old course

- T-61.5040 Learning Models and Methods
- no more lectures, last exam in March, 2008

Little overlap expected in parts with courses like

- T-61.3050 Machine Learning: Basic Principles
- T-61.5130 Machine Learning and Neural Networks
- T-61.3020 Principles of Pattern Recognition

Some overlap is good!

Resources on Machine Learning

Machine Learning: Basic Principles course book

- Ethem Alpaydin: Introduction to Machine Learning, MIT Press, 2004
- Conferences on Machine Learning:
  - European Conference on Machine Learning (ECML), co-located with the Principles of Knowledge Discovery and Data Mining (PKDD)
  - International Conference in Machine Learning (ICML), in Helsinki in July 2008, see for details: `http://icml2008.cs.helsinki.fi/`
  - Uncertainty in Artificial Intelligence (UAI), in Helsinki in July 2008, see for details: `http://uai2008.cs.helsinki.fi/`

Resources on Machine Learning

Journals in Machine Learning

- Machine Learning, Journal of Machine Learning Research, IEEE Pattern Analysis and Machine Intelligence, Pattern Recognition, Pattern Recognition Letters, Neural Computing, Neural Computation, and many others
- Also domain-related journals: BMC Bioinformatics, Bioinformatics, etc.

Community-based resources

- Mailing lists: UAI, connectionists, ML-news, ml-list, kdnuggets, etc.
- `http://en.wikipedia.org/wiki/Machine_learning`

What is machine learning?

- Machine learning people develop algorithms for computers to learn from data.
- We don't cover all of machine learning!
- The modern approach to machine learning: the probabilistic approach
- The probabilistic approach to machine learning
  - Generative models, Finite mixture models
  - Graphical models, Bayesian networks
  - Inference and learning
  - Expectation Maximization algorithm

Topics covered on the course

Central topics

- Random variables
- Independence and conditional independence
- Bayes's rule
- Naive Bayes classifier, finite mixture models, k-means clustering
- Expectation Maximization algorithm for inference and learning
- Computational algorithms for exact inference
- Computational algorithms for approximate inference
- Sampling techniques
- Bayesian modeling

Three simple examples

- Simple coin tossing with one coin
- A game two players: coin tossing with two coins
- Naive Bayes classification in a bioinformatics application

Simple coin tossing with one coin

- Throw a coin
- The coin lands either on heads (H) or tails (T).
- We don't know the outcome before the experiment
- We model the outcome with a random variable X
- $X = \{H, T\}, P(X = H) =?, P(X = T) = 1-?$

- Perform an experiment, estimate the "?"
- Parameterization: $P(X = T) = \theta, P(X = H) = 1 - \theta$
- Fixed parameters tell about the properties of the coin

Simple coin tossing with one coin

After the experiment, we have $X_1 = x_1, \ldots, X_{12} = x_{12}$

- The likelihood function is the probability of observed data $P(x_1, \ldots, x_{12};\ \theta_1, \theta_2, \ldots, \theta_{12})$
- What can we assume? What do we want to assume? Fair coin?
- Coin tosses are independent and identically distributed random variables
- Likelihood function factorizes to $P(x_1;\ \theta)P(x_2;\ \theta)\ldots P(x_{12};\theta)$
- Maximum likelihood estimator gives a parameter value that maximizes the likelihood

Guessing game with two coins

Description of the game:

- Player one, player two
- Coin number one: $P(X_1 = T) = \theta_1$ (unknown)
- Coin number two: $P(X_2 = T) = \theta_2$ (unknown)
- Player one chooses a coin randomly, either one or two
- model the choice as a random variable
- Choose coin: $P(C = c_1) = \pi_1$, or $P(C = c_2) = \pi_2$
- $\pi_1 + \pi_2 = 1 \Rightarrow \pi_2 = 1 - \pi_1$

Guessing game with two coins

We would like to do better that guessing, let's model the situation

- Outcome of the coin from coin j: $P(X|C=j)$
- Ingredients: $P(X|C=1), P(X|C=2), P(C)$
- First, the coin is chosen (secretly), then, thrown
- The outcome of the coin *depends* on the choice
- $P(X,C) = P(C)P(X|C)$
- $P(X) = \sum_{j=1}^{2} P(C=j)P(X|C=j)$

What is the probability of heads?

Guessing game with two coins

Guess which coin it was?

- $P(C = j|X)$? We know $P(C), P(X|C), P(X)$
- Use the Bayes's rule!

$$P(C|X) = \frac{P(C)P(X|C)}{P(X)}$$

Which coin was it more probably if you observed heads?

Naive Bayes classification

Classify gastric cancers using DNA copy number amplification data $X_1, \ldots, X_6$

- The observed data: $X_i = \{0, 1\}, i = 1, \ldots, 6$
- Class labels: $C = 1, 2$
- The joint probability distribution $P(X_1, X_2, X_3, X_4, X_5, X_6, C)$
- Assumptions creep in...
- X_i and X_j are conditionally independent given C
- $P(X_1, X_2, X_3, X_4, X_5, X_6, C) = P(C)P(X_1|C)P(X_2|C) \ldots P(X_6|C)$
- Interest in $P(C|X_1, X_2, \ldots, X_6)$

Demo here!

Problem sessions

Schedule for the problem sessions:

- First Problem session: 25 of January, 10.15-12.00
- Problems posted on the Web site one week before the session