Introduction to Probabilistic Models for Machine Learning

Gaurav Sharma

University of Rochester

Motivation

- What are probabilistic models and why do we want to study them?
  - Uncertainty is inherent in day to day phenomena. Examples?
  - Probability provides a framework for modeling uncertainty. Examples?
  - And, for making inferences and decisions in the face of many inter-related stochastic dependencies
- Can you think of some examples?
  - Day-to-day: Travel time from origin to destination
  - Science/Engineering applications ...

Motivation II

- Probabilistic models are used in a wide variety of fields and applications. Examples?
- Machine learning (speech and language understanding, natural language processing, clustering, . . .)
  - Computer Vision/Image Processing: segmentation, denoising, object detection/tracking, . . .
  - Communications: error correction coding, synchronization, ...
  - Bioinformatics: structure prediction for proteins, noncoding RNAs, . . .
- What are some alternative frameworks to Probabilistic models?
  - Rule based reasoning, fuzzy logic
- Why are probabilistic models often preferred?
  - Reasoning coherently in the face of uncertainty and noisy observations
  - Combining diverse sources of information in a principled fashion

Course Objective

The course aims at providing attendees a foundation in inference and estimation using probabilistic models. Starting from the broad base of probabilistic inference and estimation, the course develops the treatment of specific techniques that underlie many current day machine learning and inference algorithms. Topics covered include a review of concepts from probability and stochastic processes, IID and Markov processes, basics of inference and estimation, Maximum Aposteriori Probability (MAP) and Maximum Likelihood (ML), expectation maximization for ML estimation, hidden Markov models, stochastic context free grammars, and Markov and conditional random fields. The pedagogical approach is to illustrate the use of models via concrete examples : each model is introduced via a detailed toy example and then illustrated via one or two actual application examples.

Course Topics

- Review of Basics:
  - Basic probability concepts, probability spaces and events, independence and conditional independence, Bayes rule
  - discrete and continuous probability distributions, joint, conditional, and marginal distributions
- Probabilistic Models:
  - IID models, Mixture models, Markov Chains and Processes.
  - Hidden Markov Models (HMMs), Stochastic Context Free Grammars, Markov/Conditional Random Fields
- Inference and Parameter Estimation
  - Clustering using mixture models, Expectation Maximization (EM), Viterbi and Forward-Backward recursions for Hidden Markov Models, Inside-Out and Cocke-Younger-Kasami (CYK) algorithms for Stochastic Context Free Grammars.
- Dynamic Programming and Belief Propagation
  - Dynamic programming and Belief Propagation as generalized abstractions for common algorithms, Implementation issues: scaling and computational scheduling options.
- Applications and Approximation
  - Several drawn from research

Exercise 1: Introductions and Course Expectations

- My Background
  - Professor, ECE Dept, Univ of Rochester (UR)
  - Distinguished Researcher, Center of Excellence in Data Science, UR
  - Experience across Industry and Academia
    - Ten years at Xerox Corporation (Principal Scientist/Member Research Staff)
    - Former Director, Center for Emerging and Innovative Sciences (CEIS)
- Editor-in-Chief, 2018-2020, IEEE Transaction on Image Processing
- Editor-in-Chief, 2011-2015: Journal of Electronic Imaging (SPIE/IS&T)
- Web: `http://www.ece.rochester.edu/~gsharma`
- Broad research interests in data analytics, computer vision/image processing, color imaging, signal processing, multimedia-security, communications, bioinformatics

Research Interests

- Medical and Bio Informatics
  - Wearable sensor health analytics, RNA Secondary Structure Prediction, Flow Cytometry, multimodal signal analysis
- Imaging systems and color science
  - Color Imaging, Digital halftoning, performance evaluation and design of imaging systems, image restoration, multiprimary displays
- Multimedia security
  - Forensic privacy, watermarking, steganography, steganalysis, image/video authentication, collusion resilient fingerprinting
- Digital image and video processing/Computer vision
  - Multi-camera sensor networks, distributed estimation and coding, sensor scheduling and optimization in visual sensor networks, wide-area motion imagery

My Expectations

- Pace will be rapid
  - Please ask questions when something is unclear
- Questions are welcome
  - On all topics and all directions
  - Tangential/course unrelated questions are best for breaks
- Participation improves the course for everyone

Course Logistics and Structure

- Lecture/discussion will cover core and provide overview of additional topics
  - Prioritize understanding over coverage
  - Foundation should allow you to build upon further
  - Notes will sometimes have more material than will be actively presented
- Programming based assignments (e.g. MATLAB™/Python) for further exploring subset of concepts covered
- An extensive reading list is provided
  - Covers the course and goes beyond

Course Logistics Continued

- Course WeChat group
  - Will be used for course related announcements, posting of materials, etc
- Course Teaching Assistants
  - Two HIT PhD students are assisting as teaching assistants (TAs)
    - Yanqin Zhang (张砚钦)
    - Hong Chu (洪楚)
  - Will be assisting with lecture notes, assignments, and other materials on WeChat, etc.
  - Can also respond to questions regarding assignments, set-up, etc

Course Logistics Continued

- Short break about midway through each 4 hour session
  - Ask if you'd like another brief stretch break
- Will adapt schedule as required
- In case of any unanticipated technical glitches/network disconnects
  - Please stay connected to the Zoom session and check course WeChat group for any announcements
- Note that I am geographically located in US East coast time zone
  - Availability will depend on local time
    - 12 hours behind China Standard Time
    - Best times: before class and evening 8pm onward China Time

Pre-requisites

- Probability (will do a quick review)
  - Restrict attention to discrete "time" models, often with finite state space
    - Eliminates measure theoretic technicalities, while conveying essence
- Basic linear algebra (matrices and vectors)

Summary

- Probabilistic models are powerful tools that provide a foundational framework for many machine learning algorithms. These have typically been rediscovered in multiple application settings. The course provides a "constructionist" introduction to probabilistic models through examples and sampling of applications that help intuitive understanding by building abstraction on top of concrete examples.

References I

[1] C.M. Bishop. *Pattern recognition and machine learning*. New York, NY: Springer, 2006.

[2] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second. New York, NY: Springer-Verlag, 2009.

[3] Gareth James et al. *An Introduction to Statistical Learning, with Applications in R*. New York, NY: Springer-Verlag, 2013.

[4] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. Cambridge, MA: MIT press, 2009.