Course Recap: Introduction to Probabilistic Models for Inference and Estimation

Gaurav Sharma

University of Rochester

Course Objective

- Provide an introduction to probabilistic models through a sampling of applications. An "engineering" introduction that builds toward abstraction through concrete examples
- Constructionist rather than abstract approach
  - Detailed toy example construction
    - Complete "soup-to-nuts" description
  - Abstraction to general case
  - Research application examples
    - Realistic application scenarios
- Aim was to have cohesive and self-contained coverage
  - Illustrate clear progression of ideas and inter-linkages between topics

Course Summary/Recap I

- Review Probability Basics
  - Probability spaces: Events, outcomes, probability measure
  - Random variables and processes
- Key ideas
  - Marginalization (Marginal from joint pdf/pmf)
  - Independence (Meaning?)
  - Bayes rule and conditional probability
    - Conditional independence (Meaning?)

Course Summary/Recap II

- Elementary probabilistic models
  - IID (Independent identically distributed) model
    - No dependence
    - Characterized by pmf
  - Markov model
    - Sequential dependence: future is conditionally independent of past given present
    - Consequence: factorization of joint pmf into initial and transition probabilities
    - Characterized by state transition probabilities and initial state pmf
    - Classification of Markov chains
    - Focus: Irreducible and ergodic Markov chains

Course Summary/Recap III

- Probabilistic Inference
  - Optimal Decision Rule
    - Minimum probability of error
    - $\equiv$ maximum a posteriori probability (MAP) rule

$$\hat{i} = \arg\max_{i} p\,(i \mid \mathrm{x})$$

    - Intuitive: choose most likely decision given the data
    - A posteriori probability factors into likelihood and prior

$$p\,(i \mid \mathrm{x}) \equiv p\,(\mathrm{x} \mid i)\, p(i)$$

    - MAP reduces to maximum likelihood (ML) for equiprobable prior

Course Summary/Recap IV

- Probabilistic Parameter Estimation
  - Formulated in analogous manner to inference
  - MAP parameter estimation

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} p\left(\boldsymbol{\theta} \mid \mathrm{x}\right) \equiv \arg\max_{\boldsymbol{\theta}} p\left(\mathrm{x} \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)$$

- Intuitive: choose most likely value of parameters given the data
- A posteriori probability factors into likelihood and prior

$$p\left(\boldsymbol{\theta} \mid \mathrm{x}\right) \equiv p\left(\mathrm{x} \mid \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)$$

- MAP reduces to maximum likelihood (ML) for equiprobable prior

Course Summary/Recap V

- Parameter estimation for IID and Markov models
  - Estimators were simple and intuitive
  - Probabilities estimated as corresponding empirically observed fractions of occurrences
    - of events for IID models
    - of transitions for Markov models

Course Summary/Recap VI

- Expectation Maximization (EM)
  - Modeling using unobserved latent variables
  - More general and useful models
    - for realistic situations where there may be incomplete/missing data
  - EM for IID latent variables
    - Two coin toy example
    - Application example: WAMI vehicle detection to road map alignment
  - Gaussian mixture models (GMMs)
    - Soft clustering, EM for parameter estimation
    - Application example: Flow cytometry data analysis, rare sub-population detection
  - Convergence behavior

Course Summary/Recap VII

- Hidden Markov Models (HMMs)
  - Alternative conceptual bases:
    - Hide part of Markov chain
    - Add memory to IID latent variable models
  - Alternative representations
  - Formulation as observed process (HMM) $\equiv$ stochastic function of unobserved Markov state process
  - Characterization: state transition probabilities, symbol emission probabilities, initial state probabilities
  - "Four coin" toy example

Course Summary/Recap VIII

- Inference and estimation for Hidden Markov Models (HMMs)
  - Three basic problems: likelihood computation, most likely state sequence, parameter estimation
  - Forward recursion (likelihood), Viterbi algorithm (optimal state sequence), Forward-Backward recursion (Baum-Welch parameter re-estimation)
  - Development for toy model
  - Trellis diagram for visualization, formal algebraic derivation/justification
    - Marginalization and Bayes rule + Markov property
- HMM Applications
  - Viterbi and Symbol MAPP decoding for convolutional codes over binary symmetric channel
  - String edit distance and (RNA) sequence alignment

Course Summary/Recap IX

- Stochastic Context Free Grammars (SCFGs)
  - Generalize dependency structure of Markov chains
  - Enable efficient modeling of some long range dependencies
    - Comprehends nesting dependencies of arbitrary depth
    - Origins in language modeling/representation
  - Three basic problems: likelihood computation, most likely parse tree, parameter estimation
  - Inside (outward) computation (likelihood), CYK algorithm (optimal parse tree), Inside-Outside recursion (parameter re-estimation)
  - Application to biological sequences (RNA secondary structure)
    - Toy Nussinov recursion dynamic program motivated inside-outside structure
    - Corresponding SFCG illustrated probabilistic inference
- Strong analogy with HMMs but greater conceptual & computational complexity

Course Summary/Recap X

- Markov random fields (MRFs)
  - Neighborhood based conditional independence generalization of Markov chains
  - "Sequential" decomposition is not feasible
  - Only iterative numerical/computational methods for approximate inference/estimation
  - Key computational enabler: Hammersly-Clifford theorem
    - Characterization in terms of clique potentials
    - Connection between energy and probabilitistic models
  - Application to image segmentation

Course Summary/Recap XI

- Key algorithmic ideas/concepts
  - Dynamic Programming
    - Reuse solutions for smaller problems in solving larger problem
    - Breaks tyranny of exponential increase in complexity
    - Problem needs right structure for "search space" and "cost function"
  - Belief Propagation
    - Iterative updates of probabilities of interdependent variables
    - Yields exact marginal probabilities for variables of interest for directed acyclic graphs (DAGs)
    - Forward-Backward and Inside-Out algorithms for HMMs and SCFGs can be cast as instances of belief propagation
  - Dynamic programming and Belief Propagation as generalized abstractions for algorithms invented before/independently of formalization of these concepts
- Implementation issues: scaling and computational scheduling options

Continuing Beyond This Course I

- We discussed generative models
  - Joint distribution $p(x, z)$ of observations x and "states" z
    - Intuitive with recursion terms defined as appropriate probabilities
    - Model allows for generation of pairs x, z according to $p(x, z)$
- Discriminative models offer an alternative
  - Conditional random fields (CRFs): model condl. PMF $p(z \mid x)$
  - adequate for most inference tasks
  - require potentially fewer assumptions on dependencies with x and between elements of x and z
  - Linear chain/tree-structured CRFs are the discriminative analogs of HMMs/SCFGs
  - Algorithmic approaches generalize straightforwardly
    - Less intuitive with recursion terms corresponding to factor graph representation for conditional PMF

Continuing Beyond This Course II

- Applications
  - Natural language processing and biomolecular sequence and structure modeling using HMMs and SCFGs
- Approximation algorithms
  - Many problems of current research interest are computationally intractable
  - Good practical algorithms exist, however, for many of these computationally intractable problems
    - Often explainable as approximations to exact optimal inference/parameter estimation
    - Examples: Decoding of Turbo/LDPC Codes as an instance of Belief Propagation
    - Can also motivate applications in new domains: Turbo-Decoding of RNA Structural Alignments
- Lot of excellent resources available online
  - Tutorials from Machine Learning/Signal Processing conferences
- You will learn the material best by implementing/using the models in an application of your own
- Think creatively to incorporate models where appropriate in your work

Reminders

- Course evaluations
  - Please complete the course evaluations
  - Written feedback and suggestions, particularly welcome
  - Will help shape future editions of the course
  - Also, feel free to email me your thoughts/suggestions

Acknowledgments

- Thanks to several of my students and collaborators
  - Research examples
  - Slides
- Thank you for attending and participating