1: Introduction

Instructor: Yizhou Sun

yzsun@cs.ucla.edu

January 7, 2019
Instructor

• Yizhou Sun
  • yzsun@cs.ucla.edu
  • http://web.cs.ucla.edu/~yzsun/

• Research areas
  • Social/information network mining, graph mining, text mining, web mining
  • Data mining, machine learning
Logistics of the Course

• Grading
  • Participation: 5%
  • Homework: 25%
  • Paper presentation: 30%
    • Group-based
  • Course project: 40%
    • Group-based
Lectures

- Part I: Lectures by the instructor (5 weeks)
  - Cover the basic materials
- Part II: paper presentation by students (4-5 weeks)
  - Extended materials, which require in-depth reading of papers
- Part III: course project presentation (final week)
Homework

• 2-3 homework in Part I
• A quick quiz style homework for each paper, due every week in Part II
Paper Presentation

- **What to present**
  - Each student sign-up for one group of research papers
    - Every group can be signed by at most 4 students
- **How long for each presentation?**
  - 1 lecture, including Q&A
- **When to present**
  - From Week 6 to Week 10
- **How to present**
  - Make slides, when necessary, using blackboard
- **What else?**
  - Design a homework with 3 quick questions
    - Could be multi-choice, true/false, or other types of questions
  - Send the slides and homework (with correct answer) to me the day before the lecture
Course Project

• Research project
  • Goal: design a probabilistic graphical model to solve the candidate problems, and write a report that is potentially submitted to some venue for publication

• Teamwork
  • 3-4 people per group

• Timeline
  • Team formation due date: Week 2
  • Proposal due date: Week 5
  • Presentation due date: 3/20/2019, final exam time
  • Final report due date: 3/20/2019
    • What to submit: project report and code
Content

• What are probabilistic models
• What are structured data
• Applications
• Key tasks and challenges
A Typical Machine Learning Problem

• Given a feature vector $\mathbf{x}$, predict its label $y$ (discrete or continuous)

\[ y = f(\mathbf{x}) \]

• Example: Text classification
  
  • Given a news article, which category does it belong to?

Argentina played to a frustrating 1-1 ties against Iceland on Saturday. A stubborn Icelandic defense was increasingly tough to penetrate, and a Lionel MESSI missed penalty was a huge turning point in the match, because it likely would’ve given Argentina three points.
Probabilistic Models

- Data: \( D = \{(x_i, y_i)\}_{i=1}^{n} \)
  - \( n \): number of data points
- Model: \( p(D | \theta) \) or \( p_{\theta}(D) \)
  - Use probability distribution to address uncertainty
  - \( \theta \): parameters in the model
- Inference: ask questions about the model
  - Marginal inference: marginal probability of a variable
  - Maximum a posteriori (MAP) inference: most likely assignment of variables
- Learning: learn the best parameters \( \theta \)
The I.I.D. Assumption

• Assume data points are independent and identically distributed (i.i.d.)
  • \( p(D|\theta) = \prod_i p(x_i, y_i|\theta) \) (if modeling joint distribution)
  • \( p(D|\theta) = \prod_i p(y_i|x_i, \theta) \) (if modeling conditional distribution, conditional i.i.d.)

• Example: linear regression
  • \( y_i|x_i, \beta \sim N(x_i^T \beta, \sigma^2) \)
    • \( y_i = x_i^T \beta + \varepsilon_i, \) where \( \varepsilon_i \sim N(0, \sigma^2) \)

\[
p(D|\beta) = \prod_i p(y_i|x_i, \beta) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y_i - x_i^T \beta)^2}{2\sigma^2} \right\}
\]

\( L(\beta): \text{likelihood function} \)
Content

• What are probabilistic models

• What are structured data

• Applications

• Key tasks and challenges
Structured Data

- Dependency between data points
  - Dependency are described by links
- Example: paper citation network
  - Citation between papers introduces dependency
Examples of Structured Data

- Text
- sequence
- Image
  - Grid / regular graph
- Social/Information Network
  - General graph
Roles of Data Dependency

- I.I.D. or conditional I.I.D. assumption no longer holds
  - \( p(D|\theta) \neq \prod_i p(x_i, y_i|\theta), \) or
  - \( p(D|\theta) \neq \prod_i p(y_i|x_i, \theta) \)

- Example
  - In paper citation network, a paper is more likely to share the same label (research area) of its references

<table>
<thead>
<tr>
<th>Paper i’s label</th>
<th>Paper j’s label</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.4</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Suppose i cites j or j cites i
Scope of This Course

- A subset of probabilistic graphical model
  - Consider data dependency
  - Markov Random Fields, Conditional Random Fields, Factor Graph, and their applications in text, image, and social/information networks

- A full cover of probabilistic graphical models can be found:
  - Stanford course
    - Stefano Ermon, CS 228: Probabilistic Graphical Models
    - Daphne Koller, Probabilistic Graphical Models, YouTube
  - CMU course
    - Eric Xing, 10-708: Probabilistic Graphical Models
Content

• What are probabilistic models

• What are structured data

• Applications

• Key tasks and challenges
Text NER

• Named-Entity Recognition
  • Given a predefined label set, determine each word’s label
    • E.g., B-PER, I-PER, O
  • Possible solution: Conditional random field
  • https://nlp.stanford.edu/software/CRF-NER.html

Named Entity Recognition:

1. At the W party Thursday night at Chateau Marmont, Cate Blanchett barely made it up in the elevator.
Image Semantic Labeling

- Determine the label of each pixel
  - Given a predefined label set, determine each pixel’s label
  - Possible solution: Conditional random field
Social Network Node Classification

- Attribute prediction of Facebook users
  - E.g., gender
  - Zheleva et al., Higher-order Graphical Models for Classification in Social and Affiliation Networks, NIPS’2010
Content

• What are probabilistic models

• What are structured data

• Applications

• Key tasks and challenges
Key Tasks

• Model
  • From data model to graphical model
  • Define joint probability of all the data according to graphical model
    • \( p(D|\theta) \) or \( p_\theta(D) \)

• Inference
  • Marginal inference: marginal probability of a variable
  • Maximum a posteriori (MAP) inference: most likely assignment of variables

• Learning
  • Learn the best parameters \( \theta \)
Key Challenges

• Design challenges in modeling
  • How to use heuristics to design meaningful graphical model?
• Computational challenges in inference and learning
  • Usually are NP-hard problems
  • Need approximate algorithms
Course Overview

- Preliminary
  - Introduction
- Basic probabilistic models
  - Naïve Bayes
  - Logistic Regression
- Warm up: Hidden Markov Models
  - Forward Algorithm, Viterbi Algorithm, The Forward-Backward Algorithm
- Markov Random Fields
  - General MRF, Pairwise MRF
  - Variable elimination, sum-product message passing, max-product message passing, exponential family, pseudo-likelihood
- Conditional Random Fields
  - General CRF, Linear Chain CRF
- Factor Graph
Probability Review

- Follow Stanford CS229 Probability Notes
  - [http://cs229.stanford.edu/section/cs229-prob.pdf](http://cs229.stanford.edu/section/cs229-prob.pdf)
Major Concepts

• Elements of Probability
  • Sample space, event space, probability measure
  • Conditional probability
  • Independence, conditional independence

• Random variables
  • Cumulative distribution function, Probability mass function (for discrete random variable), Probability density function (for continuous random variable)
  • Expectation, variance
  • Some frequently used distributions
    • Discrete: Bernoulli, binomial, geometric, poisson
    • Continuous: uniform, exponential, normal

• More random variables
  • Joint distribution, marginal distribution, joint and marginal probability mass function, joint and marginal density function
  • Chain rule
  • Bayes’ rule
  • Independence
  • Expectation, conditional expectation, and covariance
Summary

• What are probabilistic models
  • Model uncertainty
• What are structured data
  • Use links to capture dependency between data
• Applications
  • Text, image, social/information network
• Key tasks and challenges
  • Modeling, inference, learning
References

- Zheleva et al., Higher-order Graphical Models for Classification in Social and Affiliation Networks, NIPS’2010
- https://cs.stanford.edu/~ermon/cs228/index.html
- https://nlp.stanford.edu/software/CRF-NER.html