CSci 8980: Advanced Topics in Graphical Models

Instructor: Arindam Banerjee

September 4, 2007

General Information

- Course Number: CSci 8980
- Class: Tu Th 09:45-11:00 am
- Location: 156 Amundson Hall
- Instructor: Arindam Banerjee
- Office Hours: EE/CS 6-213 Tu Th 11 am - 12 noon
- Web page: http://www-users.itlabs.umn.edu/classes/Fall-2007/csci8980-graph
- Email: banerjee@cs.umn.edu

Course Work

- Paper Reviews: 30% of total grade

Course Work

- Paper Reviews: 30% of total grade
  - Review 10 papers from the ‘Papers’ section

Course Work

- Paper Reviews: 30% of total grade
  - Review 10 papers from the ‘Papers’ section
  - Each paper is 3% of total grade

Course Work

- Paper Reviews: 30% of total grade
  - Review 10 papers from the 'Papers' section
  - Each paper is 3% of total grade
  - 1-2 page reviews, due before class

Course Work

- Paper Reviews: 30% of total grade
  - Review 10 papers from the 'Papers' section
  - Each paper is 3% of total grade
  - 1-2 page reviews, due before class
  - Cannot submit more than 10 reviews

Course Work

- Paper Reviews: 30% of total grade
  - Review 10 papers from the 'Papers' section
  - Each paper is 3% of total grade
  - 1-2 page reviews, due before class
  - Cannot submit more than 10 reviews
  - Guidelines in class page

Course Work

- Paper Reviews: 30% of total grade
  - Review 10 papers from the 'Papers' section
  - Each paper is 3% of total grade
  - 1-2 page reviews, due before class
  - Cannot submit more than 10 reviews
  - Guidelines in class page
- Paper Presentation: 15% of total grade

Course Work

- Paper Reviews: 30% of total grade
  - Review 10 papers from the 'Papers' section
  - Each paper is 3% of total grade
  - 1-2 page reviews, due before class
  - Cannot submit more than 10 reviews
  - Guidelines in class page
- Paper Presentation: 15% of total grade
  - Present one of the papers in the 'Papers' section

Course Work

- Paper Reviews: 30% of total grade
  - Review 10 papers from the 'Papers' section
  - Each paper is 3% of total grade
  - 1-2 page reviews, due before class
  - Cannot submit more than 10 reviews
  - Guidelines in class page
- Paper Presentation: 15% of total grade
  - Present one of the papers in the 'Papers' section
  - Talk should be 40-45 minutes

Course Work

- Paper Reviews: 30% of total grade
  - Review 10 papers from the 'Papers' section
  - Each paper is 3% of total grade
  - 1-2 page reviews, due before class
  - Cannot submit more than 10 reviews
  - Guidelines in class page
- Paper Presentation: 15% of total grade
  - Present one of the papers in the 'Papers' section
  - Talk should be 40-45 minutes
  - Guidelines in class page

Course Work

- Paper Reviews: 30% of total grade
  - Review 10 papers from the 'Papers' section
  - Each paper is 3% of total grade
  - 1-2 page reviews, due before class
  - Cannot submit more than 10 reviews
  - Guidelines in class page
- Paper Presentation: 15% of total grade
  - Present one of the papers in the 'Papers' section
  - Talk should be 40-45 minutes
  - Guidelines in class page
- Class Participation: 10% of total grade

Course Work

- Paper Reviews: 30% of total grade
  - Review 10 papers from the 'Papers' section
  - Each paper is 3% of total grade
  - 1-2 page reviews, due before class
  - Cannot submit more than 10 reviews
  - Guidelines in class page
- Paper Presentation: 15% of total grade
  - Present one of the papers in the 'Papers' section
  - Talk should be 40-45 minutes
  - Guidelines in class page
- Class Participation: 10% of total grade
  - Discussion, Q&A in class

Course Work

- Paper Reviews: 30% of total grade
  - Review 10 papers from the 'Papers' section
  - Each paper is 3% of total grade
  - 1-2 page reviews, due before class
  - Cannot submit more than 10 reviews
  - Guidelines in class page
- Paper Presentation: 15% of total grade
  - Present one of the papers in the 'Papers' section
  - Talk should be 40-45 minutes
  - Guidelines in class page
- Class Participation: 10% of total grade
  - Discussion, Q&A in class
  - Contributions need to be constructive/useful

Course Work (Contd.)

- Class Project: 45% of total grade

Course Work (Contd.)

- Class Project: 45% of total grade
  - Has the following components

Course Work (Contd.)

- Class Project: 45% of total grade
  - Has the following components
    - Proposal: 0% (due Sep 27)

Course Work (Contd.)

- Class Project: 45% of total grade
  - Has the following components
    - Proposal: 0% (due Sep 27)
    - Revised Proposal: 5% (due Oct 18)

Course Work (Contd.)

- Class Project: 45% of total grade
  - Has the following components
    - Proposal: 0% (due Sep 27)
    - Revised Proposal: 5% (due Oct 18)
    - Midterm Report: 25% (due Nov 15)

Course Work (Contd.)

- Class Project: 45% of total grade
  - Has the following components
    - Proposal: 0% (due Sep 27)
    - Revised Proposal: 5% (due Oct 18)
    - Midterm Report: 25% (due Nov 15)
    - Presentation: 20% (last 2 weeks of class)

Course Work (Contd.)

- Class Project: 45% of total grade
  - Has the following components
    - Proposal: 0% (due Sep 27)
    - Revised Proposal: 5% (due Oct 18)
    - Midterm Report: 25% (due Nov 15)
    - Presentation: 20% (last 2 weeks of class)
    - Final Report: 50% (due Dec 13, noon)

Course Work (Contd.)

- Class Project: 45% of total grade
  - Has the following components
    - Proposal: 0% (due Sep 27)
    - Revised Proposal: 5% (due Oct 18)
    - Midterm Report: 25% (due Nov 15)
    - Presentation: 20% (last 2 weeks of class)
    - Final Report: 50% (due Dec 13, noon)
  - Groups of 1-2 (you can work on your own)

Course Work (Contd.)

- Class Project: 45% of total grade
  - Has the following components
    - Proposal: 0% (due Sep 27)
    - Revised Proposal: 5% (due Oct 18)
    - Midterm Report: 25% (due Nov 15)
    - Presentation: 20% (last 2 weeks of class)
    - Final Report: 50% (due Dec 13, noon)
  - Groups of 1-2 (you can work on your own)
  - Stop by if you want to discuss project ideas

Course Work (Contd.)

- Class Project: 45% of total grade
  - Has the following components
    - Proposal: 0% (due Sep 27)
    - Revised Proposal: 5% (due Oct 18)
    - Midterm Report: 25% (due Nov 15)
    - Presentation: 20% (last 2 weeks of class)
    - Final Report: 50% (due Dec 13, noon)
  - Groups of 1-2 (you can work on your own)
  - Stop by if you want to discuss project ideas
  - Guidelines in class page

Topics

- Warmup

Topics

- Warmup
  - Finite Mixture Models, EM, Exponential Families

Topics

- Warmup
  - Finite Mixture Models, EM, Exponential Families
  - Aspect Models, PLSI, Latent Dirichlet Allocation

Topics

- Warmup
  - Finite Mixture Models, EM, Exponential Families
  - Aspect Models, PLSI, Latent Dirichlet Allocation
- Models

Topics

- Warmup
  - Finite Mixture Models, EM, Exponential Families
  - Aspect Models, PLSI, Latent Dirichlet Allocation
- Models
  - Dirichlet Processes

Topics

- Warmup
  - Finite Mixture Models, EM, Exponential Families
  - Aspect Models, PLSI, Latent Dirichlet Allocation
- Models
  - Dirichlet Processes
  - Infinite Mixture Models

Topics

- Warmup
  - Finite Mixture Models, EM, Exponential Families
  - Aspect Models, PLSI, Latent Dirichlet Allocation
- Models
  - Dirichlet Processes
  - Infinite Mixture Models
  - Hierarchical Models

Topics

- Warmup
  - Finite Mixture Models, EM, Exponential Families
  - Aspect Models, PLSI, Latent Dirichlet Allocation
- Models
  - Dirichlet Processes
  - Infinite Mixture Models
  - Hierarchical Models
  - Gaussian Processes

Topics

- Warmup
  - Finite Mixture Models, EM, Exponential Families
  - Aspect Models, PLSI, Latent Dirichlet Allocation
- Models
  - Dirichlet Processes
  - Infinite Mixture Models
  - Hierarchical Models
  - Gaussian Processes
- Inference

Topics

- Warmup
  - Finite Mixture Models, EM, Exponential Families
  - Aspect Models, PLSI, Latent Dirichlet Allocation
- Models
  - Dirichlet Processes
  - Infinite Mixture Models
  - Hierarchical Models
  - Gaussian Processes
- Inference
  - MCMC, Gibbs Sampling

Topics

- Warmup
  - Finite Mixture Models, EM, Exponential Families
  - Aspect Models, PLSI, Latent Dirichlet Allocation
- Models
  - Dirichlet Processes
  - Infinite Mixture Models
  - Hierarchical Models
  - Gaussian Processes
- Inference
  - MCMC, Gibbs Sampling
  - Variational Inference

Topics

- Warmup
  - Finite Mixture Models, EM, Exponential Families
  - Aspect Models, PLSI, Latent Dirichlet Allocation
- Models
  - Dirichlet Processes
  - Infinite Mixture Models
  - Hierarchical Models
  - Gaussian Processes
- Inference
  - MCMC, Gibbs Sampling
  - Variational Inference
  - Expectation Propagation

Topics

- Warmup
  - Finite Mixture Models, EM, Exponential Families
  - Aspect Models, PLSI, Latent Dirichlet Allocation
- Models
  - Dirichlet Processes
  - Infinite Mixture Models
  - Hierarchical Models
  - Gaussian Processes
- Inference
  - MCMC, Gibbs Sampling
  - Variational Inference
  - Expectation Propagation
- Applications

Topics

- Warmup
    - Finite Mixture Models, EM, Exponential Families
    - Aspect Models, PLSI, Latent Dirichlet Allocation
- Models
    - Dirichlet Processes
    - Infinite Mixture Models
    - Hierarchical Models
    - Gaussian Processes
- Inference
    - MCMC, Gibbs Sampling
    - Variational Inference
    - Expectation Propagation
- Applications
    - Topic Modeling, Social Network Analysis

Topics

- Warmup
  - Finite Mixture Models, EM, Exponential Families
  - Aspect Models, PLSI, Latent Dirichlet Allocation
- Models
  - Dirichlet Processes
  - Infinite Mixture Models
  - Hierarchical Models
  - Gaussian Processes
- Inference
  - MCMC, Gibbs Sampling
  - Variational Inference
  - Expectation Propagation
- Applications
  - Topic Modeling, Social Network Analysis
  - Computational Biology

Finite Mixture Models (FMM)

- A FMM is a probabilistic model of the form

$$p(x|\alpha, \Theta) = \sum_{h=1}^{k} \alpha_h p_h(x|\theta_h)$$

Finite Mixture Models (FMM)

- A FMM is a probabilistic model of the form

$$p(x|\alpha, \Theta) = \sum_{h=1}^{k} \alpha_h p_h(x|\theta_h)$$

- α is a discrete distribution over k components

Finite Mixture Models (FMM)

- A FMM is a probabilistic model of the form

$$p(x|\alpha, \Theta) = \sum_{h=1}^{k} \alpha_h p_h(x|\theta_h)$$

- α is a discrete distribution over k components
- $p_h(x|\theta_h)$ is the distribution of the h^{th} component

Finite Mixture Models (FMM)

- A FMM is a probabilistic model of the form

$$p(x|\alpha, \Theta) = \sum_{h=1}^{k} \alpha_h p_h(x|\theta_h)$$

- α is a discrete distribution over k components
- $p_h(x|\theta_h)$ is the distribution of the h^{th} component
- Widely used for model-based clustering

FMM (Contd.)

- Generative Model

FMM (Contd.)

- Generative Model
  - Sample $h \sim \alpha$

FMM (Contd.)

- Generative Model
  - Sample $h \sim \alpha$
  - Sample $x \sim p(x|\theta_h)$

FMM (Contd.)

- Generative Model
  - Sample $h \sim \alpha$
  - Sample $x \sim p(x|\theta_h)$
- Given a set of samples $\mathcal{X} = \{x_1, \ldots, x_n\}$

FMM (Contd.)

- Generative Model
  - Sample $h \sim \alpha$
  - Sample $x \sim p(x|\theta_h)$
- Given a set of samples $\mathcal{X} = \{x_1, \ldots, x_n\}$
  - Estimation problem: Which set of parameters are *most likely*

$$(\alpha^*, \Theta^*) = \text{argmax}_{(\alpha,\Theta)} \sum_{i=1}^{n} \log p(x_i|\alpha, \Theta)$$

FMM (Contd.)

- Generative Model
  - Sample $h \sim \alpha$
  - Sample $x \sim p(x|\theta_h)$
- Given a set of samples $\mathcal{X} = \{x_1, \ldots, x_n\}$
  - Estimation problem: Which set of parameters are *most likely*

$$(\alpha^*, \Theta^*) = \text{argmax}_{(\alpha,\Theta)} \sum_{i=1}^{n} \log p(x_i|\alpha, \Theta)$$

  - Minimizes KL-divergence to the empirical distribution

FMM (Contd.)

- Generative Model
  - Sample $h \sim \alpha$
  - Sample $x \sim p(x|\theta_h)$
- Given a set of samples $\mathcal{X} = \{x_1, \ldots, x_n\}$
  - Estimation problem: Which set of parameters are *most likely*

$$(\alpha^*, \Theta^*) = \text{argmax}_{(\alpha,\Theta)} \sum_{i=1}^{n} \log p(x_i|\alpha, \Theta)$$

  - Minimizes KL-divergence to the empirical distribution
  - Inference problem: Which component z_i generated sample x_i?

Learning Mixture Models

- Estimation: Need to maximize

$$\sum_{i=1}^{n} \log \left(\sum_{h=1}^{k} \alpha_h p_h(x_i | \theta_h) \right)$$

Learning Mixture Models

- Estimation: Need to maximize

$$\sum_{i=1}^{n} \log \left(\sum_{h=1}^{k} \alpha_h p_h(x_i | \theta_h) \right)$$

- Direct optimization w.r.t. (α, Θ) is "tricky"

Learning Mixture Models

- Estimation: Need to maximize

$$\sum_{i=1}^{n} \log \left(\sum_{h=1}^{k} \alpha_h p_h(x_i | \theta_h) \right)$$

- Direct optimization w.r.t. (α, Θ) is "tricky"
- Expectation Maximization (EM) is the standard approach

Learning Mixture Models

- Estimation: Need to maximize

$$\sum_{i=1}^{n} \log \left(\sum_{h=1}^{k} \alpha_h p_h(x_i | \theta_h) \right)$$

- Direct optimization w.r.t. (α, Θ) is "tricky"
- Expectation Maximization (EM) is the standard approach
- Recent years have seen progress on alternative methods

EM: The Basic Idea

- Let z_i be the latent component generating x_i

EM: The Basic Idea

- Let z_i be the latent component generating x_i
- The complete log-likelihood is given by

$$\log p(X, Z|\alpha, \Theta) = \sum_{i=1}^{n} \log(\alpha_{z_i} p_{z_i}(x_i|\theta_{z_i}))$$

EM: The Basic Idea

- Let z_i be the latent component generating x_i
- The complete log-likelihood is given by

$$\log p(X, Z|\alpha, \Theta) = \sum_{i=1}^{n} \log(\alpha_{z_i} p_{z_i}(x_i|\theta_{z_i}))$$

- A random variable, function of latent variables z_i

EM: The Basic Idea

- Let z_i be the latent component generating x_i
- The complete log-likelihood is given by

$$\log p(X, Z|\alpha, \Theta) = \sum_{i=1}^{n} \log(\alpha_{z_i} p_{z_i}(x_i|\theta_{z_i}))$$

- A random variable, function of latent variables z_i
- EM implements the following:

EM: The Basic Idea

- Let z_i be the latent component generating x_i
- The complete log-likelihood is given by

$$\log p(X, Z|\alpha, \Theta) = \sum_{i=1}^{n} \log(\alpha_{z_i} p_{z_i}(x_i|\theta_{z_i}))$$

- A random variable, function of latent variables z_i
- EM implements the following:
  - Inference (E-step): Obtain the distributions $p(z_i|x_i, \alpha, \Theta)$

EM: The Basic Idea

- Let z_i be the latent component generating x_i
- The complete log-likelihood is given by

$$\log p(X, Z|\alpha, \Theta) = \sum_{i=1}^{n} \log(\alpha_{z_i} p_{z_i}(x_i|\theta_{z_i}))$$

- A random variable, function of latent variables z_i
- EM implements the following:
    - Inference (E-step): Obtain the distributions $p(z_i|x_i, \alpha, \Theta)$
    - Estimation (M-step): Obtain parameters (α, Θ) that maximize

$$E_Z[\log p(X, Z|\alpha, \Theta)]$$