CS281A/Stat241A Lecture 23

Variational Methods

Peter Bartlett

Key ideas of this lecture

- Variational versus sampling methods
- Examples of algorithms:
  - Loopy belief propagation
  - Mean field algorithm
- Graphical model exponential families
  - Examples: Ising model; Gaussian MRF.
  - Mean parameters, marginal polytope.
  - Mean $\leftrightarrow$ natural parameters
  - Conjugate duality
  - Variational representation
- Mean field algorithm

Inference

Consider a graphical model (say undirected):

$$p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C).$$

- The inference problem:

Given observations x_E

of variables in an evidence set, $E \subset V$,

and a set of variables $F \subset V$,

$\ldots$ find $p(x_F | x_E = \bar{x}_E)$.

Maximizing *a posteriori* Probability

Consider a graphical model (say undirected):

$$p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C).$$

- Maximize *a posteriori* probability:

Given observations x_E

of variables in an evidence set, $E \subset V$,

$\ldots$ find $\arg\max_x p(x|x_E = \bar{x}_E)$.

Variational Methods

- Represent quantity of interest as solution to (or value of) an optimization problem.
- Then approximate the optimization problem:
  - Approximate the constraint set.
  - Approximate the criterion.

Sampling versus Variational Methods

Sampling Methods:

- Are asymptotically exact.
- But mixing can be slow.

Variational Methods:

- Are deterministic, and typically fast.
- But are approximations, and the approximation might be poor.

Key ideas of this lecture

- Variational versus sampling methods
- Examples of algorithms:
  - Loopy belief propagation
  - Mean field algorithm
- Graphical model exponential families
  - Examples: Ising model; Gaussian MRF.
  - Mean parameters, marginal polytope.
  - Mean $\leftrightarrow$ natural parameters
  - Conjugate duality
  - Variational representation
- Mean field algorithm

Examples: Loopy Belief Propagation

- Recall **Belief Propagation** for trees:
  1. Incorporate the evidence through an evidence potential:

$$\psi^E(x_i) = \begin{cases} \delta(x_i, \bar{x}_i) & \text{if } i \in E, \\ 1 & \text{otherwise.} \end{cases}$$

  2. Pass messages (potentials) along the edge from j to i of the form

$$m_{j,i}(x_i) = \sum_{x_j} \left(\psi^E(x_j)\psi(x_i, x_j) \prod_{k \in N(j)\setminus\{i\}} m_{k,j}(x_j) \right),$$

where $N(j) = \{k : \{k, j\} \in \mathcal{E}\}$.

Examples: Loopy Belief Propagation

3. Pass messages (potentials) along the edge from j to i of the form

$$m_{j,i}(x_i) = \sum_{x_j} \left(\psi^E(x_j)\psi(x_i, x_j) \prod_{k \in N(j) \setminus \{i\}} m_{k,j}(x_j) \right),$$

   where $N(j) = \{k : \{k, j\} \in \mathcal{E}\}$. This corresponds to the potential obtained from eliminating the subtree rooted at j and away from i.

4. Follow the protocol:
   Node j sends message $m_{j,i}$ to node i iff it has received all messages $m_{k,j}$ for $k \in N(j) \setminus \{i\}$.

Examples: Loopy Belief Propagation

5. Calculate

$$p(x_i|\bar{x}_E) = \frac{1}{Z}\psi^E(x_i)\prod_{k\in N(i)} m_{k,i}(x_i).$$

Examples: Loopy Belief Propagation

- Instead of the protocol:

  Node j sends message $m_{j,i}$ to node i iff it has received all messages $m_{k,j}$ for $k \in N(j) \setminus \{i\}$

  Consider:

  1. $m_{j,i}^{(0)}(x_i) = 1$ for all $\{i, j\} \in \mathcal{E}$.
  2. At iteration $t = 1, 2, \ldots$,

$$m_{j,i}^{(t)}(x_i) = \sum_{x_j} \left(\psi^E(x_j)\psi(x_i, x_j) \prod_{k \in N(j) \setminus \{i\}} m_{k,j}^{(t-1)}(x_j) \right).$$

Examples: Loopy Belief Propagation

Node j sends message $m_{j,i}$ to node i iff it has received all messages $m_{k,j}$ for $k \in N(j) \setminus \{i\}$

$$m_{j,i}^{(t)}(x_i) = \sum_{x_j} \left(\psi^E(x_j)\psi(x_i, x_j) \prod_{k \in N(j) \setminus \{i\}} m_{k,j}^{(t-1)}(x_j) \right).$$

These protocols are **equivalent** for trees:

- By induction (working inwards from the leaves), we can see that, for t at least as large as the depth of the subtree rooted at j and away from i,

$$m_{j,i}^{(t)}(x_i) = m_{j,i}(x_i).$$

Examples: Loopy Belief Propagation

1. $m_{j,i}^{(0)}(x_i) = 1$ for all $\{i, j\} \in \mathcal{E}$.
2. At iteration $t = 1, 2, \ldots$,

$$m_{j,i}^{(t)}(x_i) = \sum_{x_j} \left(\psi^E(x_j)\psi(x_i, x_j) \prod_{k \in N(j)\backslash\{i\}} m_{k,j}^{(t-1)}(x_j) \right).$$

- This protocol makes sense for arbitrary graphs: pretend that the graph is a tree.
- If there are a few long cycles, we might expect this to work well.

Key ideas of this lecture

- Variational versus sampling methods
- Examples of algorithms:
  - Loopy belief propagation
  - Mean field algorithm
- Graphical model exponential families
  - Examples: Ising model; Gaussian MRF.
  - Mean parameters, marginal polytope.
  - Mean $\leftrightarrow$ natural parameters
  - Conjugate duality
  - Variational representation
- Mean field algorithm

Examples: Mean Field Algorithm

Consider a discrete undirected model (Markov random field):

$$x_u \in \mathcal{X}_u \qquad\qquad |\mathcal{X}_u| < \infty,$$

$$\begin{aligned}
\ln \psi_{u,v}(x_u, x_v) &= \sum_{i,j} \theta_{u,i;v,j} \mathbf{1}[x_u = i]\mathbf{1}[x_v = j] \\
&= \theta_{u;v}(x_u, x_v), \\
\ln \psi_v(x_v) &= \sum_i \theta_{v,i} \mathbf{1}[x_v = i] \\
&= \theta_v(x_v).
\end{aligned}$$

$$p(x) \propto \exp\left(\sum_{v \in V} \theta_v(x_v) + \sum_{\{u,v\} \in E} \theta_{u,v}(x_u, x_v) \right).$$

Examples: Mean Field Algorithm

Consider Gibbs sampling in the Ising model, a discrete MRF with $x_v \in \{0, 1\}$:

$$X_v^{(t+1)} = \begin{cases} 1 & \text{if } U \leq 1/\left(1 + \exp\left(-\theta_v - \sum_{u \in N(v)} \theta_{v,u} X_u^{(t)}\right)\right), \\ 0 & \text{otherwise,} \end{cases}$$

where U is chosen uniformly from $[0, 1]$.

Examples: Mean Field Algorithm

- Suppose that $\sum_{u \in N(v)} \theta_{v,u} X_u^{(t)}$ is close to its expectation.
- For example, if the set $N(v)$ is large, this is true with high probability.
- Then we could replace the random $X_v^{(t)}$ values with their expectations, μ_v, to obtain

$$\mu_v := \frac{1}{1 + \exp\left(-\theta_v - \sum_{u \in N(v)} \theta_{v,u} \mu_u\right)}.$$

- This is called the naive mean field algorithm for the Ising model.
- It can also be viewed as message passing.

Issues to Consider

For message passing algorithms like loopy belief propagation or the mean field algorithm,

- Do these message passing updates have a fixed point?
- Is it (close to) the desired conditional probability?
- Do the updates converge to the fixed point?

We'll see that these algorithms can be viewed as methods for solving approximate versions of variational formulations of the inference problem.

Key ideas of this lecture

- Variational versus sampling methods
- Examples of algorithms:
  - Loopy belief propagation
  - Mean field algorithm
- Graphical model exponential families
  - Examples: Ising model; Gaussian MRF.
  - Mean parameters, marginal polytope.
  - Mean $\leftrightarrow$ natural parameters
  - Conjugate duality
  - Variational representation
- Mean field algorithm

Graphical Models: Exponential Families

Example: Ising Model $(x_v \in \{0,1\})$.

$$p(x) = \exp\left(\sum_{v\in V}\theta_v x_v + \sum_{\{u,v\}\in E}\theta_{u,v}x_u x_v - A(\theta)\right).$$

for $\theta \in \Omega = \{\theta : A(\theta) < \infty\} = \mathbb{R}^{|V|+|E|}$.

- *Regular* (Ω is open.)
- *Minimal* (no p-invariant subspace of Ω.)

Graphical Models: Exponential Families

Generalization of Ising: pairwise MRF
$(x_v \in \{0, 1, \ldots, r-1\})$.

$$p(x) = \exp\left(\sum_{v \in V} \sum_i \theta_{v,i} \mathbf{1}[x_v = i] \right.$$
$$\left. + \sum_{\{u,v\} \in E} \sum_{i,j} \theta_{u,i;v,j} \mathbf{1}[x_u = i] \mathbf{1}[x_v = j] - A(\theta)\right),$$

for $\theta \in \Omega = \{\theta : A(\theta) < \infty\} = \mathbb{R}^{r|V| + r^2|E|}$.

- Regular (Ω is open.)
- Non-minimal or *overcomplete*.

Graphical Models: Exponential Families

- Special case: Hidden Markov model with y observed.
  - $\theta_{t,i}$ corresponds to $\log p(y_t|x_t = i)$.
  - $\theta_{t,i;t+1,j}$ corresponds to $\log p(x_{t+1} = j|x_t = i)$.
- Another generalization: Higher order interactions, that is, k-cliques, with $k > 2$.

Graphical Models: Exponential Families

Example: Gaussian Markov random field.

- For an undirected graph (V, E), define the sufficient statistics x_v, x_v^2, $x_u x_v$ for $v \in V$ and $\{u, v\} \in E$.

$$p(x) = \exp\left(\langle \theta, x \rangle + \frac{1}{2} \left\langle \Theta, xx' \right\rangle - A(\theta, \Theta) \right),$$

where the second inner product is

$$\left\langle \Theta, xx' \right\rangle = \text{tr}(\Theta x x').$$

Graphical Models: Exponential Families

Example: Gaussian Markov random field.

- Here, the natural parameters are a vector $\theta \in \mathbb{R}^{|V|}$ and a symmetric positive definite matrix $\Theta \in \mathbb{R}^{|V|\times|V|}$, with $\Theta_{u,v} = 0$ if $\{u, v\} \notin E$.
- The natural parameters corresponding to x_v^2 and $x_u x_v$ correspond to the non-zero entries of the precision matrix Θ.

In this case, the parameters are restricted to

$$\Omega = \{(\theta, \Theta) : A(\theta, \Theta) < \infty\} = \{(\theta, \Theta) : \Theta < 0\},$$

where the Θ are symmetric matrices with zero entries where edges are missing.

Key ideas of this lecture

- Variational versus sampling methods
- Examples of algorithms:
  - Loopy belief propagation
  - Mean field algorithm
- Graphical model exponential families
  - Examples: Ising model; Gaussian MRF.
  - Mean parameters, marginal polytope.
  - Mean $\leftrightarrow$ natural parameters
  - Conjugate duality
  - Variational representation
- Mean field algorithm

Mean Parameters

Fix a density p defined with respect to a reference distribution h.
For a sufficient statistic ϕ_α, define the mean parameter μ_α as

$$\mu_\alpha = \mathbb{E}_p[\phi_\alpha(X)] = \int \phi_\alpha(x)p(x)h(dx).$$

For d sufficient statistics, we can define the d-vector of mean parameters, $\mu = (\mu_1, \ldots, \mu_d)$.
Define the set $\mathcal{M}$ of *realizable mean parameters* as

$$\mathcal{M} = \left\{\mu \in \mathbb{R}^d : \exists p \text{ s.t. } \forall\alpha,\ \mathbb{E}_p[\phi_\alpha(X)] = \mu_\alpha\right\}$$

if $\mathcal{X}$ is finite: $\quad = \text{co}\{\phi(x) : x \in \mathcal{X}\},$

where co represents the convex hull.

Mean Parameters: Ising Model

$$p(x) = \exp\left(\sum_{v \in V} \theta_v x_v + \sum_{\{u,v\} \in E} \theta_{u,v} x_u x_v - A(\theta) \right).$$

The vector of sufficient statistics is

$$\phi(x) = (x_v : v \in V,\ x_u x_v : \{u, v\} \in E).$$

and the mean parameters are

$$\mu_v = \Pr(X_v = 1),$$
$$\mu_{u,v} = \Pr(X_u = X_v = 1).$$

Mean Parameters: Ising Model

The vector of sufficient statistics is

$$\phi(x) = (x_v : v \in V,\ x_u x_v : \{u, v\} \in E).$$

Then $\mathcal{M}$ is the *marginal polytope*,

$$\mathcal{M} = \text{co}\{\phi(x) : x \in \{0,1\}^{|V|}\},$$

the convex hull of the sufficient statistic values.
It is the set of achievable singleton and pairwise marginal probabilities.

Mean Parameters: Gaussian MRF

$$p(x) = \exp\left(\langle\theta, x\rangle + \frac{1}{2}\left\langle\Theta, xx'\right\rangle - A(\theta, \Theta)\right).$$

The mean parameters are (μ, Σ), where

$$\mu = \mathbb{E}[X], \qquad \Sigma = \mathbb{E}[XX'].$$

Easy to check that $\Sigma - \mu\mu' \geq 0$ is necessary and sufficient. That is,

$$\mathcal{M} = \left\{(\mu, \Sigma) : \Sigma - \mu\mu' \geq 0\right\}.$$

Notice that $\mathcal{M}$ is again convex.

Key ideas of this lecture

- Variational versus sampling methods
- Examples of algorithms:
  - Loopy belief propagation
  - Mean field algorithm
- Graphical model exponential families
  - Examples: Ising model; Gaussian MRF.
  - Mean parameters, marginal polytope.
  - Mean $\leftrightarrow$ natural parameters
  - Conjugate duality
  - Variational representation
- Mean field algorithm

Mean $\leftrightarrow$ Natural Parameters

Recall:

1. $$\nabla A(\eta) = \mathbb{E}\phi(x),$$
   $$\nabla^2 A(\eta) = \text{Var}\,\phi(x).$$

2. For a regular family, the gradient mapping

   $$\nabla A : \Omega \to \mathcal{M}$$

   is one-to-one iff the representation is minimal.

Mean $\leftrightarrow$ Natural Parameters

3. The forward mapping, $\theta \mapsto \mu$, corresponds to computing expectations of sufficient statistics.
4. The reverse mapping, $\mu \mapsto \theta$, corresponds to computing a maximum likelihood estimate of θ for sample average μ.
5. The maximum entropy p satisfying a constraint on μ is in the exponential family.
   In particular, $\nabla A : \Omega \to \mathcal{M}$ is onto the interior of $\mathcal{M}$.

Key ideas of this lecture

- Variational versus sampling methods
- Examples of algorithms:
  - Loopy belief propagation
  - Mean field algorithm
- Graphical model exponential families
  - Examples: Ising model; Gaussian MRF.
  - Mean parameters, marginal polytope.
  - Mean $\leftrightarrow$ natural parameters
  - Conjugate duality
  - Variational representation
- Mean field algorithm

Conjugate Duality: Definition

Given a function $A : \Omega \to \mathbb{R}$, the conjugate dual is

$$A^*(\mu) = \sup_{\theta \in \Omega} \left(\langle \mu, \theta \rangle - A(\theta) \right),$$

where $\mu \in \mathbb{R}^d$ for $\Omega \subseteq \mathbb{R}^d$.

- A^* is convex (a maximum of linear functions).
- Think of $A^* : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$.
- If A is convex $(+ \ldots)$, $A^{**} = A$. We can think of A^* as capturing the shape of a convex A through the locations of the tangent planes to its epigraph.
- If A is log normalization for an exponential family, $\langle \mu, \theta \rangle - A(\theta)$ is (constant plus) log likelihood with sample average μ and natural parameter θ.

Conjugate Duality

Theorem:

1. For μ in the interior of $\mathcal{M}$, let $\theta(\mu)$ satisfy

$$\mathbb{E}_{p_{\theta(\mu)}}[\phi(X)] = \nabla A(\theta(\mu)) = \mu.$$

   Then

$$A^*(\mu) = -H(p_{\theta(\mu)}) = \int_{\mathcal{X}} \log p_{\theta(\mu)}(x) p_{\theta(\mu)}(x) h(dx).$$

2. For μ outside the closure of $\mathcal{M}$,

$$A^*(\mu) = \infty.$$

Conjugate Duality

Theorem:

3. For $\theta \in \Omega$, we have the **variational representation**

$$A(\theta) = \sup_{\mu \in \mathcal{M}} \left(\langle \theta, \mu \rangle - A^*(\mu) \right).$$

4. For $\theta \in \Omega$,

$$A(\theta) = \langle \theta, \mu(\theta) \rangle - A^*(\mu(\theta)),$$

where

$$\mu(\theta) := \mathbb{E}_{p_\theta}[\phi(X)] = \nabla A(\theta).$$

Conjugate Duality

- $-A^*(\mu)$ is the value of the maximum entropy problem for mean parameter μ
- $-A^*(\mu) = -\infty$ for infeasible μ.
- Forward mapping: $\nabla A : \Omega \to \mathcal{M}$.
- Backward mapping: $\nabla A^* : \text{int}(\mathcal{M}) \to \Omega$.

Conjugate Duality: Bernoulli

$$X \in \{0, 1\},$$
$$\phi(x) = x,$$
$$p(x) = \exp(\theta x - A(\theta)),$$
$$A(\theta) = \log(\exp(0) + \exp(\theta)) = \log(1 + \exp(\theta)),$$
$$\Omega = \{\theta \in \mathbb{R} : A(\theta) < \infty\} = \mathbb{R}.$$

Conjugate Duality: Bernoulli

$$A(\theta) = \log(1 + \exp(\theta)),$$

$$A^*(\mu) = \sup_{\theta \in \mathbb{R}} \left(\theta\mu - \log(1 + \exp(\theta))\right).$$

Solving for the maximizing θ gives

$$\mu = \frac{\exp(\theta)}{1 + \exp(\theta)}$$

$$\theta = \log \frac{\mu}{1-\mu} \qquad \text{for } \mu \in (0,1),$$

$$\begin{aligned} A^*(\mu) &= \mu \log \frac{\mu}{1-\mu} - \log \frac{1}{1-\mu} \\ &= \mu \log \mu + (1-\mu)\log(1-\mu) = -H(p_{\theta(\mu)}). \end{aligned}$$

Conjugate Duality: Bernoulli

And if μ is outside $[0, 1]$?

$$\frac{d}{d\theta}\mu\theta = \mu,$$

$$\frac{d}{d\theta}\log(1+\exp(\theta)) = \frac{\exp(\theta)}{1+\exp(\theta)} \in (0,1).$$

So for μ outside $[0, 1]$,

$$A^*(\mu) = \sup_{\theta\in\mathbb{R}} \left(\theta\mu - \log(1+\exp(\theta))\right) = \infty.$$

Key ideas of this lecture

- Variational versus sampling methods
- Examples of algorithms:
  - Loopy belief propagation
  - Mean field algorithm
- Graphical model exponential families
  - Examples: Ising model; Gaussian MRF.
  - Mean parameters, marginal polytope.
  - Mean $\leftrightarrow$ natural parameters
  - Conjugate duality
  - Variational representation
- Mean field algorithm

Variational Representation of $A(\theta)$

For $\theta \in \Omega$,

$$A(\theta) = \sup_{\mu \in \mathcal{M}} \left(\langle \theta, \mu \rangle - A^*(\mu) \right)$$
$$= \sup_{\mu \in \mathcal{M}} \left(\langle \theta, \mu \rangle + H(p_{\theta(\mu)}) \right).$$

- Solving this optimization problem gives the value $A(\theta)$ and the mean parameters $\mu = \mathbb{E}_\theta[\phi(X)]$.
- These correspond to the expectation of the sufficient statistics. (conditional expectation, if evidence has been incorporated into θ).
- For example, for discrete pairwise MRFs, they give the marginal singleton and pairwise distributions.

Variational Representation of $A(\theta)$

$$A(\theta) = \sup_{\mu \in \mathcal{M}} \left(\langle \theta, \mu \rangle + H(p_{\theta(\mu)}) \right).$$

- We can approximate this optimization problem to obtain a simpler problem:
  - Approximate $\mathcal{M}$ by a simpler set $\hat{\mathcal{M}}$.
    $\hat{\mathcal{M}} \subset \mathcal{M}$ gives a lower bound.
    $\mathcal{M} \subset \hat{\mathcal{M}}$ gives an upper bound.
  - Approximate $H(p_{\theta(\mu)})$ by something simpler.

Key ideas of this lecture

- Variational versus sampling methods
- Examples of algorithms:
  - Loopy belief propagation
  - Mean field algorithm
- Graphical model exponential families
  - Examples: Ising model; Gaussian MRF.
  - Mean parameters, marginal polytope.
  - Mean $\leftrightarrow$ natural parameters
  - Conjugate duality
  - Variational representation
- Mean field algorithm

Examples: Mean Field Algorithm

Consider an Ising model (binary pairwise Markov random field):

$$\begin{aligned}
x_u &\in \{0,1\}, \\
\psi_{u,v}(x_u, x_v) &= \theta_{u,v} x_u x_v, \\
\psi_v(x_v) &= \theta_v x_v. \\
p(x) &\propto \exp\left(\sum_{v \in V} \theta_v x_v + \sum_{\{u,v\} \in E} \theta_{u,v} x_u x_v \right). \\
\mu_{u,v} &= \Pr(x_u = x_v = 1), \\
\mu_v &= \Pr(x_v = 1).
\end{aligned}$$

Mean Field Algorithm

- We approximate $\mathcal{M}$ with a smaller set:

$$\hat{\mathcal{M}} = \{\mu : \mu_{u,v} = \mu_u \mu_v\}.$$

- This adds independence, so $\hat{\mathcal{M}} \subset \mathcal{M}$.
- Thus, we can represent the distribution as

$$p(x;\theta) = \prod_{v \in V} p_v(x_v;\theta).$$

- Hence, the entropy is

$$H(p_{\theta(\mu)}) = \mathbb{E} \log p(X;\theta) = \sum_{v \in V} \mathbb{E} \log p_v(X_v;\theta)$$

$$= \sum_{v \in V} (\mu_v \log \mu_v + (1 - \mu_v) \log(1 - \mu_v)).$$

Mean Field Algorithm

- So we have

$$
\begin{aligned}
A(\theta) &= \sup_{\mu \in \hat{\mathcal{M}}} \left(\langle \theta, \mu \rangle + H(p_{\theta(\mu)}) \right) \\
&= \sup_{\mu \in \hat{\mathcal{M}}} \left(\sum_{v \in V} \theta_v \mu_v + \sum_{\{u,v\} \in E} \theta_{u,v} \mu_u \mu_v \right. \\
&\qquad \left. - \sum_{v \in V} \left(\mu_v \log \mu_v + (1 - \mu_v) \log(1 - \mu_v) \right) \right) .
\end{aligned}
$$

Mean Field Algorithm

- We can solve this with coordinate maximization: Calculate gradient of criterion w.r.t. μ_v:

$$\theta_v + \sum_{u \in N(v)} \theta_{u,v}\mu_u - (1 + \log \mu_v - 1 - \log(1 - \mu_v))$$
$$= \theta_v + \sum_{u \in N(v)} \theta_{u,v}\mu_u - \log \frac{\mu_v}{1 - \mu_v}.$$

- Setting to zero gives

$$\mu_v = \frac{1}{1 + \exp\left(-\theta_v - \sum_{u \in N(v)} \theta_{u,v}\mu_u\right)},$$

which is the mean field update.

Mean Field Algorithm

Summary:

- We approximate $\mathcal{M}$ with a smaller set:

$$\hat{\mathcal{M}} = \left\{ \mu : \mu_{u,v} = \mu_u \mu_v \right\}.$$

- Solve for $A(\theta)$ and μ with coordinate maximization:

$$\mu_v := \frac{1}{1 + \exp\left(-\theta_v - \sum_{u \in N(v)} \theta_{u,v} \mu_u \right)}.$$

Key ideas of this lecture

- Variational versus sampling methods
- Examples of algorithms:
  - Loopy belief propagation
  - Mean field algorithm
- Graphical model exponential families
  - Examples: Ising model; Gaussian MRF.
  - Mean parameters, marginal polytope.
  - Mean $\leftrightarrow$ natural parameters
  - Conjugate duality
  - Variational representation
- Mean field algorithm

Announcements

- My office hours:
  Thursday Nov 19 (today), 1-2pm, in 723 SD Hall.