CS 6347

Lecture 7

Approximate MAP & Lagrange Multipliers

Reparameterization

- The messages passed in max-product can be used to construct a **reparameterization** of the joint distribution

$$p(x_1, \dots, x_n) = \frac{1}{Z} \prod_{i \in V} \phi_i(x_i) \prod_{(i,j) \in E} \psi_{ij}(x_i, x_j)$$

and

$$p(x_1, \dots, x_n) = \frac{1}{Z} \prod_{i \in V} \left[\phi_i(x_i) \prod_{k \in N(i)} m_{k \to i}(x_i) \right] \prod_{(i,j) \in E} \frac{\psi_{ij}(x_i, x_j)}{m_{i \to j}(x_j) m_{j \to i}(x_i)}$$

Reparameterization

$$p(x_1, \dots, x_n) = \frac{1}{Z} \prod_{i \in V} \left[\phi_i(x_i) \prod_{k \in N(i)} m_{k \to i}(x_i) \right] \prod_{(i,j) \in E} \frac{\psi_{ij}(x_i, x_j)}{m_{i \to j}(x_j) m_{j \to i}(x_i)}$$

- Reparameterizations do not change the partition function, the MAP solution, or the factorization of the joint distribution
  - They just push "weight" around between the different factors

Upper Bounds

$$\max_{x_1,\ldots,x_n} p(x_1,\ldots,x_n) \leq \frac{1}{Z} \prod_{i \in V} \max_{x_i} \phi_i(x_i) \prod_{(i,j) \in E} \max_{x_i,x_j} \psi_{ij}(x_i,x_j)$$

- This provides an upper bound on the optimization problem
  - Do other reparameterizations provide better bounds?

Duality

$$L(m) = \frac{1}{Z} \prod_{i \in V} \max_{x_i} \left[\phi_i(x_i) \prod_{k \in N(i)} m_{k \to i}(x_i) \right] \prod_{(i,j) \in E} \max_{x_i, x_j} \left[\frac{\psi_{ij}(x_i, x_j)}{m_{i \to j}(x_j) m_{j \to i}(x_i)} \right]$$

- We construct a ***dual*** optimization problem

$$\min_m L(m) \geq \max_x p(x)$$

- Last time, we saw how to attempt to minimize L via coordinate descent

Integer Programming

- We can express the MAP problem as a 0,1 integer programming problem
  - Convert a maximum of a product into a maximum of a sum by taking logs
  - Introduce indicator variables, τ, to represent the chosen assignment

Integer Programming

- Introduce variables
  - $\tau_i(x_i) \in \{0,1\}$ for each $i \in V$ and x_i
  - $\tau_{ij}(x_i, x_j) \in \{0,1\}$ for each $(i,j) \in E$ and x_i, x_j
- The linear objective function is then

$$\max_{\tau} \sum_{i \in V} \sum_{x_i} \tau_i(x_i) \log \phi_i(x_i) + \sum_{(i,j) \in E} \sum_{x_i, x_j} \tau_{ij}(x_i, x_j) \log \psi_{ij}(x_i, x_j)$$

where the τ's are required to satisfy certain marginalization conditions

Integer Programming

$$\max_{\tau} \sum_{i \in V} \sum_{x_i} \tau_i(x_i) \log \phi_i(x_i) + \sum_{(i,j) \in E} \sum_{x_i, x_j} \tau_{ij}(x_i, x_j) \log \psi_{ij}(x_i, x_j)$$

such that

$$\sum_{x_i} \tau_i(x_i) = 1$$

For all $i \in V$

$$\sum_{x_j} \tau_{ij}(x_i, x_j) = \tau_i(x_i)$$

For all $(i, j) \in E, x_i$

$$\tau_i(x_i) \in \{0,1\}$$

For all $i \in V, x_i$

$$\tau_{ij}(x_i, x_j) \in \{0,1\}$$

For all $(i, j) \in E, x_i, x_j$

Integer Programming

$$\max_{\tau} \sum_{i \in V} \sum_{x_i} \tau_i(x_i) \log \phi_i(x_i) + \sum_{(i,j) \in E} \sum_{x_i, x_j} \tau_{ij}(x_i, x_j) \log \psi_{ij}(x_i, x_j)$$

such that

These constraints define the vertices of the **marginal polytope** (set of all valid marginal distributions)

$$\sum_{x_i} \tau_i(x_i) = 1 \qquad \text{For all } i \in V$$

$$\sum_{x_j} \tau_{ij}(x_i, x_j) = \tau_i(x_i) \qquad \text{For all } (i,j) \in E, x_i$$

$$\tau_i(x_i) \in \{0,1\} \qquad \text{For all } i \in V, x_i$$

$$\tau_{ij}(x_i, x_j) \in \{0,1\} \qquad \text{For all } (i,j) \in E, x_i, x_j$$

Linear Relaxation

- The integer program can be relaxed into a linear program by replacing the 0,1 integrality constraints with linear constraints
  - This relaxed set of constraints forms the **local marginal polytope**
    - The τ's no longer correspond to an achievable marginal distribution, so we call them pseudo-marginals
  - We call it a relaxation because the constraints have been relaxed: all solutions to the IP are contained as solutions of the LP
- Linear programming problems can be solved in polynomial time

Linear Relaxation

$$\max_{\tau} \sum_{i \in V} \sum_{x_i} \tau_i(x_i) \log \phi_i(x_i) + \sum_{(i,j) \in E} \sum_{x_i, x_j} \tau_{ij}(x_i, x_j) \log \psi_{ij}(x_i, x_j)$$

such that

$\sum_{x_i} \tau_i(x_i) = 1$	For all $i \in V$
$\sum_{x_j} \tau_{ij}(x_i, x_j) = \tau_i(x_i)$	For all $(i, j) \in E, x_i$
$\tau_i(x_i) \in [0,1]$	For all $i \in V, x_i$
$\tau_{ij}(x_i, x_j) \in [0,1]$	For all $(i, j) \in E, x_i, x_j$

LP vs. Dual

- Both the LP relaxation and the dual $L(m)$ provide an upper bound on the MAP objective function
  - That is, finding an optimal collection of messages is equivalent to finding the best pseudo-marginals
- In fact, they are equivalent optimization problems: this seems quite surprising because the problems look so different
  - The proof uses the method of Lagrange multipliers (a standard mathematical technique to construct dual optimization problems)

General Optimization

$$\min_{x \in \mathbb{R}^n} f_0(x)$$

subject to:

$$\begin{aligned} f_i(x) &\leq 0, \qquad i = 1, \ldots, m \\ h_i(x) &= 0, \qquad i = 1, \ldots, p \end{aligned}$$

Lagrangian

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- Incorporate constraints into a new objective function
- λ and ν are vectors of ***Lagrange multipliers***
- The Lagrange multipliers can be thought of as soft constraints

Duality

- Construct a dual function by minimizing the Lagrangian over the primal variables

$$g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu)$$

- $g(\lambda, \nu) = -\infty$ whenever the Lagrangian is not bounded from below

Duality

- The dual has a number of important features
  - It is always concave (even if the primal is not)
  - $g(\lambda, \nu) \leq L(x, \lambda, \nu)$ for all $x \in D$
  - $L(x, \lambda, \nu) \leq f_0(x)$ for any feasible x, $\lambda \geq 0$
    - x is feasible if it satisfies all of the constraints
    - Proof?
  - This gives, $g(\lambda, \nu) \leq f_0(x)$ for all feasible x, $\lambda \geq 0$, ν

Duality

- As before, we can find the best lower bound

$$\sup_{\lambda \geq 0, \nu} g(\lambda, \nu)$$

  - This is called the dual problem
  - The pair (λ, ν) is called dual-feasible if $\lambda \geq 0$ and $g(\lambda, \nu) > -\infty$

Some Examples

- Minimize $x^2 + y^2$ subject to $x + y \geq 2$

- Minimize $x^2 + y^2$ subject to $x + y \leq 3$

- Maximize $-x \log x - y \log y - z \log z$ subject to $x, y, z \geq 0$ and $x + y + z = 1$

Weak & Strong Duality

- In general, the optimal value of the dual is always a lower bound on the optimal value of the primal
  - This property is called weak duality
- When the optimal values agree, we call it strong duality

Weak & Strong Duality

- The difference between the primal optimal value and the dual optimal value is called the duality gap

- Certain conditions on the constraints and the objective function guarantee strong duality

Tightness of the MAP LP

- When is it that solving the MAP LP (or equivalently, the dual optimization) is the same as solving the integer programming problem?
  - We say that there is no duality gap (or that the dual is tight) when this is the case
  - The answer can be expressed as a structural property of the graph