CS 228T: HW 4

Marco Cusumano-Towner

May 1, 2012

1 EP for TrueSkill model

1.1 EP

TODO

1.2 Gibbs sampling

$$\begin{aligned}P(T_k|\mathbf{y},\mathbf{w},\mathbf{T}_{-k}) &= P(T_k|y_k,w_{k1},w_{k2})\\ &\propto N(T_k;w_{k1}-w_{k2},pv)\mathbf{1}\left[\operatorname{sign}(T_k)=y_k\right]\end{aligned}$$

Let v be the prior variance.

$$\begin{aligned}P(w_i|\mathbf{w}_{-i},\mathbf{T},\mathbf{y}) &\propto P(w_i)\prod_k P(T_k|w_i,w_{k2})^{\mathbf{1}[(k,1)=i]}P(T_k|w_{k1},w_i)^{\mathbf{1}[(k,2)=i]}\\ &= P(w_i)\prod_{k\in G_{i1}}N(T_k;w_i-w_{k2},1)\prod_{k\in G_{i2}}N(T_k;w_{k2}-w_i,1)\end{aligned}$$

The product of these Gaussians is another Gaussian. We derive the new mean and variance μ_i and σ_i^2. Focusing on the exponential term, we add the contributions from all the Gaussians:

$$\exp\left[-\frac{1}{2}\left[\frac{1}{v}w_i^2+\sum_{k\in G_{i1}}(T_k-(w_i-w_{k2}))^2+\sum_{k\in G_{i2}}(T_k-(w_{k1}-w_i))^2\right]\right]$$

$$\exp\left[-\frac{1}{2}\left[\frac{1}{v}w_i^2+\sum_{k\in G_{i1}}((T_k+w_{k2})-w_i)^2+\sum_{k\in G_{i2}}((T_k-w_{k1})+w_i)^2\right]\right]$$

$$\exp\left[-\frac{1}{2}\left[a_iw_i^2+b_iw_i+c_i\right]\right]$$

where

$$a_i = \frac{1}{v} + N_i$$

$$b_i = \sum_{k \in G_{i1}} -2(T_k + w_{k2}) + \sum_{k \in G_{i2}} 2(T_k - w_{k1})$$

$$c_i = \sum_{k \in G_{i1}} (T_k + w_{k2})^2 + \sum_{k \in G_{i2}} (T_k - w_{k1})^2$$

Dividing inside the exponential by a_i gives

$$\exp\left[-\frac{1}{2}\left[w_i^2 + \frac{b_i}{a_i} w_i + \frac{c_i}{a_i}\right] a_i\right]$$

Comparing with the standard form for the Gaussian $(x - \mu)^2 = x^2 - 2\mu x + \mu^2$, we set $-2\mu_i = \frac{b_i}{a_i}$, giving

$$\mu_i = -\frac{1}{2}\frac{b_i}{a_i} = \frac{\sum_{k \in G_{i1}} (w_{k2} + T_k) + \sum_{k \in G_{i2}} (w_{k1} - T_k)}{\frac{1}{v} + N_i}$$

And the new variance is given by

$$\sigma_i^2 = \frac{1}{a_i} = \frac{1}{\frac{1}{v} + N_i}$$

The μ_i and the σ_i define the Gaussian that we sample from to get a new w_i sample.

2 Approximating the marginal polytope

2.1 Show that for any clique tree, $L(T) = M(T)$

$L(T) \subseteq M(T)$ because for a tree, calibrated beliefs define a reparameterized distribution where $P(X_i) = \beta_i(C_i)$ (Theorem 10.4)
$M(T) \subseteq L(T)$ because any distribution over T can be represented as calibrated beliefs.

2.2 Give counterexample for $L(G) \neq M(G)$

2.3 cycle inequalities

(a) Consider a cycle C. Start at a node X_A, with assignment x_A, and traverse the cycle, keeping track of the 'current' assignment y. The current assignment is the assignment to the variable you just landed on. Initialize $y = x_A$. Every time you pass a cut edge, you flip the bit of the current assignment. The cycle will reach back to X_A. If there were an odd number of cuts, the current assignment $y \neq x_A$, contradiction.

(b) First, note that this quantity must be ≥ 0, since it is the sum of indicator functions. Now, we show that it is odd. Together these imply the quantity is ≥ 1.

$$\begin{aligned}
\sum_{(i,j)\in C-F} \mathbf{1}\left[x_i \neq x_j\right] + \sum_{(i,j)\in F} \mathbf{1}\left[x_i = x_j\right] &= \sum_{(i,j)\in C} \mathbf{1}\left[x_i \neq x_j\right] - \sum_{(i,j)\in F} \mathbf{1}\left[x_i \neq x_j\right] + |F| - \sum_{(i,j)\in F} \mathbf{1}\left[x_i \neq x_j\right] \\
&= \sum_{(i,j)\in C} \mathbf{1}\left[x_i \neq x_j\right] - 2\sum_{(i,j)\in F} \mathbf{1}\left[x_i \neq x_j\right] + |F| \\
&= \text{ even } - \text{ even } + \text{ odd} \\
&= \text{ odd}
\end{aligned}$$

(c) Taking the expecation of this quantity with respect to Q gives

$$E_Q\left[\sum_{(i,j)\in C-F} \mathbf{1}\left[x_i \neq x_j\right] + \sum_{(i,j)\in F} \mathbf{1}\left[x_i = x_j\right]\right] \geq 1 \quad \forall C, F : |F| \text{ odd}$$

$$\sum_{(i,j)\in C-F} \beta_{ij}(0,1) + \beta_{ij}(1,0) + \sum_{(i,j)\in F} \beta_{ij}(0,0) + \beta_{ij}(1,1) \geq 1 \quad \forall C, F : |F| \text{ odd}$$

3 Region graphs and generalized belief propagation

1. No it's not valid. the center variable x_{22} does not have a single bottom 'sink' region, since it is included in all four of the pairwise regions. Therefore, we add a region consisting of $\{x_{22}\}$ with edges from the four pairwise regions. To satisfy the constraints on the κ_r, we set κ_r for this new region to $+1$.
2. Introducing Lagrange multipliers $\{\lambda_r\} \cup \{\lambda_{s\to r, c_r}\}$, and defining $c_{s\setminus r}$ to be an assignment to the variables in C_s that are not in C_r for each $s \to r$ relationship, the Lagrangian is:

$$\begin{aligned}
L = &\sum_r \kappa_r \sum_{c_r} \beta_r(c_r) \log \psi_r(c_r) - \sum_r \kappa_r \beta_r(c_r) \log \beta_r(c_r) \\
&- \sum_r \lambda_r (\sum_{c_r} \beta_r(c_r) - 1) - \sum_{s\to r} \sum_{c_r} \lambda_{s\to r, c_r} (\sum_{c_{s\setminus r}} \beta_s(c_r, c_{s\setminus r}) - \beta_r(c_r))
\end{aligned}$$

Taking derivatives with respect to a $\beta_r(c_r)$, we get terms from the objective as well as terms corresponding to $s \to r$ relationships and terms corresponding to $r \to s$ relationships:

$$\frac{\partial L}{\partial \beta_r(c_r)} = \kappa_r \left(\log \psi_r(c_r) - (1 + \log \beta_r(c_r))\right) - \lambda_r + \sum_{s\to r} \lambda_{s\to r, c_r} - \sum_{r\to s} \lambda_{r\to s, c_{s:r}}$$

where $c_{s:r}$ denotes the (unique) setting of c_s that agrees with c_r for some $r \to s$ relationship. Setting the derivative equal to zero and re-organizing gives a fixed-point equation:

$$\kappa_r \left(\log \psi_r(c_r) - 1 - \log \beta_r(c_r)\right) = \lambda_r - \sum_{s \to r} \lambda_{s \to r, c_r} + \sum_{r \to s} \lambda_{r \to s, c_{s:r}}$$

$$\log \beta_r(c_r) = \frac{-\lambda_r + \sum_{s \to r} \lambda_{s \to r, c_r} - \sum_{r \to s} \lambda_{r \to s, c_{s:r}}}{\kappa_r} + \log \psi_r(c_r) - 1$$

$$\beta_r(c_r) = \exp(-1) \exp(\frac{-\lambda_r}{\kappa_r}) \psi_r(c_r) \prod_{s \to r} \exp(\frac{\lambda_{s \to r, c_r}}{\kappa_r}) \prod_{r \to s} \exp(\frac{-\lambda_{r \to s, c_{s:r}}}{\kappa_r})$$

$$\beta_r(c_r) = \exp(-1) \exp(\frac{-\lambda_r}{\kappa_r}) \psi_r(c_r) \frac{\prod_{s \to r} \exp(\frac{\lambda_{s \to r, c_r}}{\kappa_r})}{\prod_{r \to s} \exp(\frac{\lambda_{r \to s, c_{s:r}}}{\kappa_r})}$$

4 Exponential families and the marginal polytope

4.1 Show that M is convex

Suppose $\mu_1, \mu_2 \in M$. Then for some p_1 and p_2, we have $\mu_1 = E_{p_1}[\tau(x)]$ and $\mu_2 = Ep_2[\tau(x)]$. Let $\mu_3 = \beta\mu_1 + (1-\beta)\mu_2$ for $0 \leq \beta \leq 1$. Then

$$\mu_3 = \int_x \left(\beta p_1(x) + (1-\beta) p_2(x)\right) \tau(x)$$

It suffices to show that $p_3 = \beta p_1 + (1-\beta) p_2$ is a valid probability distribution, because then $\mu_3 = E_{p_3}[\tau(x)]$. That p_3 is non-negative follows immediately from the fact that p_1 and p_2 are non-negative. That p_3 sums (or integrates—there is no difference in the argument) to 1 follows from:

$$\sum_x \beta p_1(x) + (1-\beta) p_2(x) = \beta \sum_x p_1(x) + \sum_x p_2(x) - \beta \sum_x p_2(x) = \beta + 1 - \beta = 1$$

4.2 Suppose χ is finite. Show that M is the convex hull of ...

The convex hull of $\{\tau(x) | x \in \chi\}$ is

$$C = \left\{ \sum_{x \in \chi} \beta_x \tau(x) \middle| \beta \geq 0, \sum_x \beta_x = 1 \right\}$$

where β_x are the coefficients in the convex combination. The set conditional on the right side is exactly the same requirement as β being a valid probability distribution, in which case the left side is the definition of expectation under β. Therefore,

$$C = \{E_\beta[\tau(x)] | \beta \text{ valid distribution } \} = M$$

4.3 Explain why M reduces to the marginal polytope for ... class of models

? what's the definition of the marginal polytope. I thought M was the marginal polytope. See Jordan Wainwright.