Chapter 19

stevenjin8

June 21, 2021

Comments and Proofs

I found this chapter a bit tough due to the sudden introduction of the node potential ϕ_t in equation 19.24. It really just means that we assign potentials both edges and nodes (unlike what section 19.3 suggests).

Exercises

Exercise 1

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\theta}_c} \log Z(\boldsymbol{\theta}) &= \frac{1}{Z(\boldsymbol{\theta})} \frac{\partial}{\partial \boldsymbol{\theta}_c} \left[\sum_{\mathbf{y}} \exp\left(\boldsymbol{\theta}_c^T \boldsymbol{\phi}_c(\mathbf{y})\right) \right] \\ &= \sum_{\mathbf{y}} \boldsymbol{\phi}_c(\mathbf{y}) \frac{1}{Z(\boldsymbol{\theta})} \exp\left(\boldsymbol{\theta}_c^T \boldsymbol{\phi}_c(\mathbf{y})\right) \\ &= \sum_{\mathbf{y}} \boldsymbol{\phi}_c(\mathbf{y}) p(\mathbf{y}|\boldsymbol{\theta})\end{aligned}$$

The cost of training an MRF is $O(rk(N+c)) = O(r(kN+kc))$. Looking at equation 19.41, we see that that we have to compute each feature for each data point per iteration. However, we only have to compute the marginals once per feature.

The cost of training a CRF $O(rNk(1+c)) = O(rNkc)$. The key difference is that the unclamped term is now conditioned on each data point. Thus, we must calculate the marginals N times (assuming that each data point is different). In other words, we must compute marginals for each iteration, data point, and feature.

Exercise 4

$$
\begin{aligned}
p(x_i = 1|\mathbf{x}_{nb_i}) &= \frac{p(x_i = 1, \mathbf{x}_{nb_i}|\boldsymbol{\theta})}{p(x_i = 0, \mathbf{x}_{nb_i}|\boldsymbol{\theta}) + p(x_i = 1, \mathbf{x}_{nb_i}|\boldsymbol{\theta})} \\
&= \frac{e_i^z}{1 + e_i^z} \\
&= \frac{1}{1 + e^{-z_i}}
\end{aligned}
$$

where

$$
\begin{aligned}
z_i &= p(x_i = 1, \mathbf{x}_{nb_i}|\boldsymbol{\theta}) \\
&= \frac{1}{Z(\boldsymbol{\theta})} \exp(h_i) \prod_{j \in nb_i} \exp(J_{ij} x_j) \\
&= \exp\left(\log Z(\boldsymbol{\theta}) + h_i + \sum_{j \in nb_i} J_{ij} x_j \right)
\end{aligned}
$$

If we keep equation 19.125 intact, but switch $x_i \in \{-1, 1\}$, then we have

$$
p(x_i = 1|\mathbf{x}_{nb_i}, \boldsymbol{\theta}) = \frac{z_i^+}{z_i^- + z_i^+} \tag{1}
$$

where z_i^+ is the unormalized probability $\tilde{p}(x_i = 1|\mathbf{x}_{nb_i}, \boldsymbol{\theta})$ and, similarly, $z_i^- = \tilde{p}(x_i = -1|\mathbf{x}_{nb_i}, \boldsymbol{\theta})$. Since

$$
\begin{aligned}
z_i^+ &= \exp\left(h_i + \sum_{j \in nb_i} J_{ij} x_j \right) \\
z_i^- &= \exp\left(-h_i - \sum_{j \in nb_i} J_{ij} x_j \right),
\end{aligned}
$$

we can see that $z_i^- = \frac{1}{z_i^-}$. Plugging back into equation 1, we get

$$
\begin{aligned}
p(x_i = 1|\mathbf{x}_{nb_i}, \boldsymbol{\theta}) &= \frac{z_i^+}{z_i^- + z_i^+} \\
&= \frac{1}{1 + (z_i^-)^2}.
\end{aligned}
$$