TTIC 31230, Fundamentals of Deep Learning

David McAllester, Autumn 2020

Exponential Softmax Backpropagation:

The Model Marginals

Notation

x is an input (e.g. an image).

$\hat{\mathcal{Y}}[N]$ is any label for x — a vector $\hat{\mathcal{Y}}[0], \ldots, \hat{\mathcal{Y}}[N-1]$ with $\hat{\mathcal{Y}}[n]$ an integer in $\{1, \ldots, L\}$.

For example n might range over the pixels of an image and $\hat{\mathcal{Y}}[n]$ names a semantic label of pixel n.

$\mathcal{Y}[N]$ is the gold label for input x — the structured label assigned to x in the training data.

Back-Propagation Through Exponential Softmax

We have node and edge score tensors computed by a deep network.

$$s^N[N, L] = f^N_\Phi(x) \qquad s^E[E, L, L] = f^E_\Phi(x)$$

$$s(\hat{\mathcal{Y}}) = \sum_n s^N[n, \hat{\mathcal{Y}}[n]] + \sum_{\langle n,m \rangle \in \mathrm{Edges}} s^E[\langle n, m \rangle, \hat{\mathcal{Y}}[n],\ \hat{\mathcal{Y}}[m]]$$

$$P_s(\hat{\mathcal{Y}}) = \operatorname*{softmax}_{\hat{\mathcal{Y}}}\ s(\hat{\mathcal{Y}}) \quad \text{all possible } \hat{\mathcal{Y}}$$

$$\mathcal{L} = -\ln P_s(\mathcal{Y}) \quad \text{gold label } \mathcal{Y}$$

For SGD we want to compute $s^N.\mathrm{grad}[N, L]$ and $s^E.\mathrm{grad}[E, L, L]$.

Model Marginals Theorem

Theorem:

$$s^N.\text{grad}[n,\ell] = P_{\hat{\mathcal{Y}}\sim P_s}(\;\hat{\mathcal{Y}}[n]=\ell\;)$$
$$-\mathbf{1}[\;\mathcal{Y}[n]=\ell\;]$$

$$s^E.\text{grad}[\langle n,m\rangle,\ell,\ell'] = P_{\hat{\mathcal{Y}}\sim P_s}(\;\hat{\mathcal{Y}}[n]=\ell\;\wedge\;\hat{\mathcal{Y}}[m]=\ell'\;)$$
$$-\mathbf{1}[\;\mathcal{Y}[n]=\ell\;\wedge\;\mathcal{Y}[m]=\ell'\;]$$

We need to compute (or approximate) the model marginals.

Proof of Model Marginals Theorem

We consider the case of node marginals, the case of edge marginals is similar.

$$
\begin{aligned}
s^N.\mathrm{grad}[n,\ell] &= \partial \mathcal{L}(\Phi, x, \mathcal{Y}) \;/\; \partial s^N[n,\ell] \\
&= \partial \left(-\ln \frac{1}{Z} \exp(s(\mathcal{Y})) \right) \;/\; \partial s^N[n,\ell] \\
&= \partial (\ln Z - s(\mathcal{Y})) \;/\; \partial s^N[n,\ell] \\
&= \left(\frac{1}{Z} \sum_{\hat{\mathcal{Y}}} e^{s(\hat{\mathcal{Y}})} \left(\partial s(\hat{\mathcal{Y}}) / \partial s^N[n,\ell] \right) \right) - \left(\partial s(\mathcal{Y}) / \partial s^N[n,\ell] \right)
\end{aligned}
$$

Proof of Model Marginals Theorem

$$s^N.\mathrm{grad}[n,\ell] = \left(\frac{1}{Z}\sum_{\hat{\mathcal{Y}}} e^{s(\hat{\mathcal{Y}})}\left(\partial s(\hat{\mathcal{Y}})/\partial s^N[n,\ell]\right)\right) - \left(\partial s(\mathcal{Y})/\partial s^N[n,\ell]\right)$$

$$= \left(\sum_{\hat{\mathcal{Y}}} P_s(\hat{\mathcal{Y}})\left(\partial s(\hat{\mathcal{Y}})/\partial s^N[n,\ell]\right)\right) - \left(\partial s(\mathcal{Y})/\partial s^N[n,\ell]\right)$$

$$s(\hat{\mathcal{Y}}) = \sum_n s^N[n,\hat{\mathcal{Y}}[n]] + \sum_{\langle n,m\rangle\in\mathrm{Edges}} s^E[\langle n,m\rangle,\hat{\mathcal{Y}}[n],\ \hat{\mathcal{Y}}[m]]$$

$$\frac{\partial s(\hat{\mathcal{Y}})}{\partial s^N[n,\ell]} = \mathbf{1}[\hat{\mathcal{Y}}[n]=\ell]$$

Proof of Model Marginals Theorem

$$s^N.\text{grad}[n,\ell] = \left(\frac{1}{Z}\sum_{\hat{\mathcal{Y}}} e^{s(\hat{\mathcal{Y}})}\left(\partial s(\hat{\mathcal{Y}})/\partial s^N[n,\ell]\right)\right) - \left(\partial s(\mathcal{Y})/\partial s^N[n,\ell]\right)$$

$$\left(\sum_{\hat{\mathcal{Y}}} P_s(\hat{\mathcal{Y}})\left(\partial s(\hat{\mathcal{Y}})/\partial s^N[n,\ell]\right)\right) - \left(\partial s(\mathcal{Y})/\partial s^N[n,\ell]\right)$$

$$= E_{\hat{\mathcal{Y}}\sim P_s}\mathbf{1}[\hat{\mathcal{Y}}[n]=\ell] - \mathbf{1}[\mathcal{Y}[n]=\ell]$$

$$= P_{\hat{\mathcal{Y}}\sim P_s}(\hat{\mathcal{Y}}[n]=\ell) - \mathbf{1}[\mathcal{Y}[n]=\ell]$$

Model Marginals Theorem

Theorem:

$$s^N.\mathrm{grad}[n,\ell] \;=\; P_{\hat{\mathcal{Y}}\sim P_s}(\;\hat{\mathcal{Y}}[n]=\ell\;)$$
$$-\mathbf{1}[\;\mathcal{Y}[n]=\ell\;]$$

$$s^E.\mathrm{grad}[\langle n,m\rangle,\ell,\ell'] \;=\; P_{\hat{\mathcal{Y}}\sim P_s}(\;\hat{\mathcal{Y}}[n]=\ell\;\wedge\;\hat{\mathcal{Y}}[m]=\ell'\;)$$
$$-\mathbf{1}[\;\mathcal{Y}[n]=\ell\;\wedge\;\mathcal{Y}[m]=\ell'\;]$$

END