TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2020

Noise Contrastive Estimation

Noise Contrastive Estimation

Gutmann and Hyvärinen, 2010

$$\Psi^* = \operatorname*{argmin}_{\Psi} E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_\Phi^{(N)}} \; -\ln P_\Psi(i|y_1,\ldots,y_N)$$

p_Φ is fixed "noise"

Assume p_Φ is both samplable and computable — we can sample from p_Φ and for any given y we can compute $p_\Phi(y)$.

Assume $P_\Psi(i|y_1,\ldots,y_N) = \mathrm{softmax}_i \, s_\Psi(y_i)$

Assume Ψ universal

Noise Contrastive Estimation

$$\Psi^* = \operatorname*{argmin}_{\Psi} E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_\Phi^{(N)}} \; -\ln P_\Psi(i|y_1,\ldots,y_N)$$

$$p_\Phi \text{ is fixed "noise"}$$

Theorem: $\mathrm{pop}(y) = \mathrm{softmax}_y \quad s_{\Psi^*}(y) + \ln p_\Phi(y)$

We then have a computable score function (energy function) for the population. We do not have the partition function Z.

Noise Contrastive Estimation

$$\Psi^* = \operatorname*{argmin}_{\Psi} E_{(i,y_1,\ldots,y_N)\sim\tilde{p}_{\Phi}^{(N)}} \; -\ln P_{\Psi}(i|y_1,\ldots,y_N)$$

p_Φ is fixed "noise"

Lemma: $P_{\Psi^*}(i|y_1,\ldots,y_N) = \mathrm{softmax}_i \;\; \ln \frac{\mathrm{pop}(y_i)}{p_\Phi(y_i)}$

Lemma Proof

$$\tilde{p}_{\Phi}^{(N)}(i \text{ and } y_1, \ldots, y_N) = \frac{1}{N}\, \mathrm{pop}(y_i) \prod_{j \neq i} p_{\Phi}(y_j)$$

$$= \alpha \frac{\mathrm{pop}(y_i)}{p_{\Phi}(y_i)}, \quad \alpha = \frac{1}{N} \prod_i p_{\Phi}(y_i)$$

$$\tilde{p}_{\Phi}^{(N)}(i \mid y_1, \ldots y_N) = \frac{\tilde{p}_{\Phi}^{(N)}(i \text{ and } y_1, \ldots, y_N)}{\sum_i \tilde{p}_{\Phi}^{(N)}(i \text{ and } y_1, \ldots, y_N)} = \frac{1}{Z} \frac{\mathrm{pop}(y_i)}{p_{\Phi}(y_i)}$$

$$= \operatorname*{softmax}_i \left(\ln \frac{\mathrm{pop}(y_i)}{p_{\Phi}(y_i)} \right)$$

Theorem Proof

$$\operatorname*{softmax}_i s_{\Psi^*}(y_i) = \operatorname*{softmax}_i \ln \frac{\mathrm{pop}(y_i)}{p_\Phi(y_i)}$$

is solved by

$$s_{\Psi^*}(y) = \ln \frac{\mathrm{pop}(y)}{p_\Phi(y)} + \ln Z$$

giving

$$\mathrm{pop}(y) = \frac{1}{Z}\exp(s_\Psi(y) + \ln p_\Phi(y))$$

END