Homework 6.1

Deep Learning 2024 Spring

Due on 2024/4/29

1 Q&A

Problem 1. (Noise Contrastive Estimation) Suppose we are using NCE to train a language model. Denote the context as h, the target word as w, and noise word samples as $\bar{\mathsf{w}}$. $\tilde{p}_{\mathsf{w}|\mathsf{h}}(w|h)$ and $p^{\theta}_{\mathsf{w}|\mathsf{h}}(w|h)$ are the target word distribution under context h of the corpus and the learning model respectively. $q_{\bar{\mathsf{w}}}(\bar{w})$ is the noise distribution introduced by NCE. Assume the learning distribution is self-normalized, i.e.,

$$p^{\theta}_{\mathsf{w}|\mathsf{h}}(w|h) = \frac{u^{\theta}(w,h)}{Z^{\theta}_{h}} \approx u^{\theta}(w,h).$$

If we choose 1 positive sample and k negative samples, the loss of NCE is given by

$$L^{k}_{\mathrm{NCE}}(\theta;h) = \sum_{w} \tilde{p}_{\mathsf{w}|\mathsf{h}}(w|h) \log\left(\frac{u^{\theta}(w,h)}{u^{\theta}(w,h) + kq_{\bar{\mathsf{w}}}(w)}\right) + \sum_{1\le i\le k,\bar{w}} q_{\bar{\mathsf{w}}}(\bar{w}) \log\left(\frac{kq_{\bar{\mathsf{w}}}(\bar{w})}{u^{\theta}(\bar{w},h) + kq_{\bar{\mathsf{w}}}(\bar{w})}\right)$$

Prove that as $k \to \infty$, $\nabla L^{k}_{\mathrm{NCE}}(\theta;h) \to \nabla L_{\mathrm{MLE}}(\theta;h)$, where

$$\nabla L_{\mathrm{MLE}}(\theta;h) = \sum_{w} \left(\tilde{p}_{\mathsf{w}|\mathsf{h}}(w|h) - p^{\theta}_{\mathsf{w}|\mathsf{h}}(w|h)\right) \nabla \log u^{\theta}(w,h).$$