TTIC 31230 Fundamentals of Deep Learning

RL Problems.

Problem 1. Consider training machine translation on a corpus of translation pairs (x, y) where x is, say, an English sentence $x_1, \ldots, \mathrm{EOS}$ and y is a French sentence $y_1, \ldots, \mathrm{EOS}$ where EOS is the "end of sentence" tag.

Suppose that we have a parameterized model defining $P_\Phi(y_t|x, y_1, \ldots, y_{t-1})$ so that $P_\Phi(y_1, \ldots, y_T|x) = \prod_{t=1}^{T'} P_\Phi(y_t|x, y_1, \ldots, y_{t-1})$ where y_T is EOS.

For a sample $\hat{y}$ from $P_\Phi(y|x)$ we also have a non-differentiable BLEU score $\mathrm{BLUE}(haty, y) \geq 0$ that is not computed until the entire output y is complete and which we would like to maximize.

(a) Give the SGD update equations for the parameters Φ for the REINFORCE algorithm for maximizing $E_{\hat{y}\sim P_\Phi(y|x)}$ for this problem.

(b) Suppose that somehow we reach a parameter setting Φ where $P_\Phi(y|x)$ assigns probability close enough to 1 for a particular translation $\hat{y}$ that in practice we will always sample the same $\hat{y}$. Suppose that this translation $\hat{y}$ has less than optimal BLEU score. Can the REINFORCE algorithm recover from this situation and consider other translations? Explain your answer.

(c) Repeat part (b) but under the assumption that $\mathrm{BLEU}(\hat{y}, y) \leq 0$ (if there is a maximum reward $R_{\max}$ we can replace R by $R - R_{\max}$).

(d) Modify the REINFORCE update equations to use a value function approximation $V_\Phi(x)$ to reduce the variance in the gradient samples. Your equations should include updates to train $V_\Phi(x)$ to predict $E_{\hat{y}\sim P(y|x)}\ \mathrm{BLEU}(\hat{y}, y)$. (Replace the reward by the "advantage" of the particular translation).