

TTIC 31230 Fundamentals of Deep Learning

Problems For Language Modeling, Translation and Attention.

Problem 1. Consider a bidirectional RNN run on a sequence of words w_1, \dots, w_T such that for each time t we have a forward hidden state $\vec{h}[t, J]$ computed from w_1, \dots, w_t and a backward hidden state $\overleftarrow{h}[t, J]$ computed from w_T, w_{T-1}, \dots, w_t .

(a) Given an explicit index (Einstein notation) definition of a cross entropy loss \mathcal{L}_t for $P(w[t] \mid w_1, \dots, w_{t-1}, w_{t+1}, \dots, w_T)$ as a function of $\vec{h}[t-1, J]$ and $\overleftarrow{h}[t+1, J]$. You should define the probability with a softmax and assume that softmax is given as a primitive. Assume a word embedding matrix $e[W, J]$ where $e[w, J]$ is the embedding vector for word w .

(b) Suppose we take the loss of a given model on a sentence w_1, \dots, w_t to be $\sum_t \mathcal{L}_t$ for \mathcal{L}_t defined as in part (a). What is the order of run time, as a function of sentence length, for the backpropagation with this loss function? Explain your answer.