ITCS 6101/8101 Homework 6: Theory (50 points)

April 7, 2026

1 Seq2Seq Modeling with Attention (25 points)

Suppose we have an RNN or LSTM encoder-decoder model with attention. The encoder processes a source sentence of 4 tokens and produces hidden states $\mathbf{h}_1 = [1, 0, 1, 0]^T$, $\mathbf{h}_2 = [0, 1, 0, 1]^T$, $\mathbf{h}_3 = [0, 0, 1, 1]^T$, and $\mathbf{h}_4 = [1, 1, 0, 0]^T$. The current decoder hidden state is $\mathbf{s}_t = [1, 1, 1, 1]^T$.

1. Use dot-product attention to compute the attention vector $\mathbf{a}_t$. Show your work.

2. Assume that the RNN/LSTM uses a binary classification head with parameters $\mathbf{w} = [-1, 0, -1, 0, 1, 1, 1, 1$ and $b = 0.1$. This classification head takes as input the concatenated vectors $[\mathbf{s}_t|\mathbf{a}_t]$. What is the label predicted at step t? Show your work.

2 Modeling Preferences (25 points)

In the Bradley-Terry Model, given two LLM outputs o_i and o_j, with associated scores z_i and z_j, the probability asociated with output i being better than output j is defined as:

$$p(o_i \succ o_j) = \sigma(z_i - z_j) \tag{1}$$

where σ is the logistic sigmoid. Prove that the difference in scores corresponds to the log-odds ratio, namely that:

$$z_i - z_j = \log \frac{p(o_i \succ o_j)}{p(o_j \succ o_i)} \tag{2}$$

3 Submission

Submit your theory responses on Canvas as one file named `hw06-theory.pdf`. It is important that you show clearly all the derivation steps. We recommend using an editor such as Overleaf for Latex, Word, or Jupyter-Notebook that allows proper formatting of equations. Alternatively, if you choose to write your solutions on paper, submit an electronic scan / photo of it exported to **PDF**. Make sure that your writing is **legible** and the scan has good quality (we will not grade solutions that we struggle to read).