Week 10 Practice

CSCI 567 Machine Learning

Fall 2025

Instructor: Haipeng Luo

1 RNNs and Transformers

1.1 Consider the following mini RNN (a picture taken from Lecture 9).

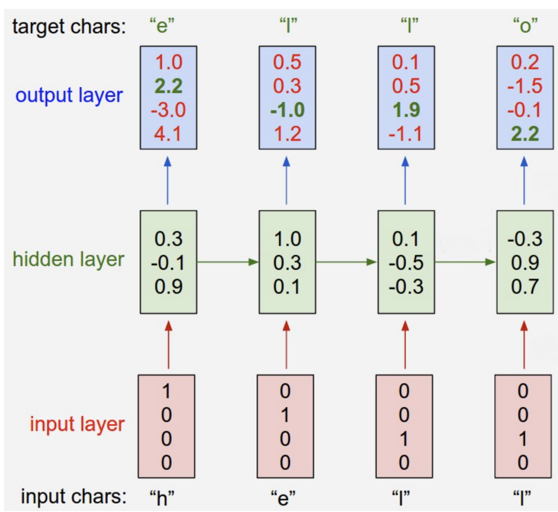

Figure 1: A mini RNN

(1) How many parameters are there in this RNN? Show your calculation and feel free to ignore all bias terms.

Since the input/output dimension is 4 and the hidden state dimension is 3, the two matrices $\boldsymbol{W}$ and $\boldsymbol{U}$ used for the state update $\boldsymbol{h}' = \sigma(\boldsymbol{W}\boldsymbol{h} + \boldsymbol{U}\boldsymbol{x})$ are 3×3 and 3×4 respectively, while the matrix $\boldsymbol{V}$ used for the output $\boldsymbol{V}\boldsymbol{h}'$ is 4×3. Therefore, in total there are $9 + 12 + 12 = 33$ parameters.

(2) What is the total cross entropy loss of the first two outputs in Figure 1? Write down you answer directly using exponentials; no need to further simplify or approximate them.

The total cross entropy loss is

$$-\ln\left(\frac{e^{2.2}}{e^{1.0} + e^{2.2} + e^{-3.0} + e^{4.1}}\right) - \ln\left(\frac{e^{-1.0}}{e^{0.5} + e^{0.3} + e^{-1.0} + e^{1.2}}\right)$$

1.2 Consider a self-attention head in an encoder of a transformer.

(1) For a 3-token input, the table below shows some partial values of the 3×3 matrix obtained after applying softmax to the attention score matrix, that is, softmax $\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\top}}{\sqrt{d_k}}\right)$. Fill in the missing values in this table (no reasoning needed).

0.4	0.4	0.2
0.1	0.3	0.6
0.3	0.6	0.1

(2) Continuing from the last question, suppose that the value vectors for these 3 input tokens are:

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

respectively. What is the corresponding output for the 2nd token? Show your calculation.

The output for the 2nd token should be the weighted sum of the value vectors according to the 2nd row of the matrix from the last question, that is:

$$0.1 \times \begin{pmatrix} 1 \\ 2 \end{pmatrix} + 0.3 \times \begin{pmatrix} -1 \\ 0 \end{pmatrix} + 0.6 \times \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0.4 \\ -0.4 \end{pmatrix}$$

2 Multi-Armed Bandits

There are 3 restaurants that are close to Alice's office, and she is having difficulty deciding which one to go for lunch each day. After using the UCB algorithm to help pick the restaurant for her for a couple weeks (see HW4), Alice is unsatisfied with its performance and feeling a lot of "regret". She suspects that this is because the quality of the meals in these restaurants does not follow a fixed distribution. Therefore, she decides to switch to the Exp3 algorithm with learning rate $\eta = 0.1$, but she needs your help to implement this idea.

2.1 Describe in one sentence how Alice should pick the restaurant on the first day.
She should pick the three restaurants uniformly at random.

2.2 Suppose that Alice ends up picking the first restaurant on the first day and really likes it (that is, reward is 1, the maximum possible reward). Describe how Alice should pick the restaurant on the second day and explain why.

Alice should still pick the restaurant uniformly at random on the second day. This is because we should use losses instead of rewards in Exp3 to ensure proper exploration. In other words, we should first convert the reward of 1 for the first restaurant to a loss of 0, which then leads to an estimated loss of 0 for all 3 restaurants. After applying softmax, this would result in a uniform distribution again.

2.3 Suppose that Alice ends up picking the second restaurant on the second day and really dislikes it (that is, reward is 0, the minimum possible reward). Describe how Alice should pick the restaurant on the third day and explain why.

Again, we first convert the reward of 0 to a loss of 1.
Since the probability of picking the second restaurant on the second day was $1/3$, the importance weighted loss estimator for the three restaurants should be $0, 3$, and 0 respectively.
Multiplying this with $-\eta = -0.1$ and further applying softmax gives the final distribution

$$\left(\frac{e^0}{e^0 + e^{-0.3} + e^0}, \frac{e^{-0.3}}{e^0 + e^{-0.3} + e^0}, \frac{e^0}{e^0 + e^{-0.3} + e^0}\right),$$

which is what Alice should sample the restaurant from on the third day.