Homework 5

Fall 2019
(Due: Dec. 6th, 2019)

Introduction

This assignment is on recurrent and recursive neural networks, which is covered in Chapter 10 of the recommended text.

Exercises

1. (Time Delay Neural Networks vs. Recurrent Neural Networks)

   Please discuss the difference between a recurrent neural network and using convolutional layers to capture temporal patterns in time-domain signals, as in Time-Delay Networks.

2. (Backpropagation through Time)

   Consider the simple RNN structure with loss in Figure 1. The network's formal description is as follows:

   $$\boldsymbol{h}^{(t)} = \tanh\left(\boldsymbol{W}\boldsymbol{h}^{(t-1)} + \boldsymbol{U}\boldsymbol{x}^{(t)} + \boldsymbol{b}\right) \tag{1}$$

   $$\boldsymbol{o}^{(t)} = \boldsymbol{V}\boldsymbol{h}^{(t)} + \boldsymbol{c} \tag{2}$$

   where $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$ are weight matrices for input-to-hidden, hidden-to-output and hidden-to-hidden connections, and the vectors $\boldsymbol{b}$ and $\boldsymbol{c}$ are biases. Before feeding the output $\boldsymbol{o}^{(t)}$ to the loss, we first need to compute the network's output vector:

   $$\hat{\boldsymbol{y}}^{(t)} = \text{Softmax}(\boldsymbol{o}^{(t)}) \in \mathbb{R}^{\tau} \tag{3}$$

   Given a sequence of input-target value pairs $\{(\boldsymbol{x}_1, \boldsymbol{y}_1), ..., (\boldsymbol{x}_\tau, \boldsymbol{y}_\tau)\}$, we have the loss $\mathcal{L}(\{(\boldsymbol{x}_t, \boldsymbol{y}_t)\}_{t=1}^{\tau}, \{\hat{\boldsymbol{y}}_t\}_{t=1}^{\tau})$.

   Your task is to explain how to compute the gradient of the loss with respect to the weight matrices, following the method of backpropagation through time (BPTT).

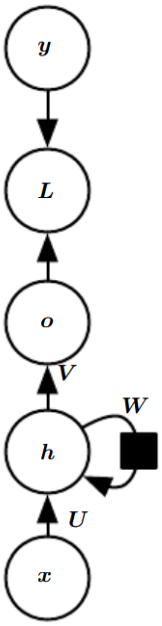

Figure 1: RNN with hidden-to-hidden recurrence.

3. (Training with Teacher Forcing)

   (i) Please explain the difference between the method's training and test setups. Why should the two be different? Feel free to use diagrams to illustrate your points.

   (ii) What problem could one encounter when we need to later use the network in an open-loop mode (network output fed back as input)? Can you provide an example solution for the problem?

   (iii) How is teacher forcing related to the maximum likelihood criterion?

4. (RNN Design)

   (i) What assumptions does an RNN make on the conditional distribution over the hidden units (regarded as random variables) at time step $t+1$ given the hidden units at t?

   Hint: Parameter sharing over time.

   (ii) Discuss three mechanisms for the RNN to determine the length of the sequence of output samples from the model.

5. (Long Short-Term Memory)

   Please describe the LSTM unit with a block diagram, along with rigorous descriptions of the forget, input and output gates. You can introduce notation for needed equations.

6. (Bidirectional RNN and Recursive Nets)

   (a) Describe the main motivation for studying and using bidirectional RNNs.

   (b) Describe an example architecture of a bidirectional RNN. Feel free to use drawings to aid your description.

   (c) Give an example of a recursive neural network. Feel free to use drawings to aid your description.

   (d) What are the advantages that recursive neural networks have over RNNs?