Recurrent Neural Networks - Tutorial

Matthias Rowold

January 19, 2023

1 Written Exercises

1.1 Exercise 1

The following recurrent neural network is given:

$$h_t = \tanh(Wh_{t-1} + Ux_t)$$
$$y_t = Vh_t$$

- Draw the computation graph for two time steps with outputs $\mathbf{y}_1$, $\mathbf{y}_2$. We assume all variables to be scalar.

- Mark input variables and parameters in circles and computations in rectangles.

- Indicate where h_1 and h_2 are computed in the graph.

1.2 Exercise 2

Let $z = \prod_i y_i$ and $\tau_i = Wh_{i-1} + Ux_i$.

- Add z and τ_i to your computation graph of Exercise 1.

- Derive the gradient $\frac{\partial z}{\partial W}$ as a function of $y_1, y_2, h_0, h_1, \tau_1, \tau_2, W, V$.
  Hint: $\frac{\partial \tanh(x)}{\partial x} = 1 - \tanh(x)^2$

- Mark the variables that have to be stored in memory (additionaly to the parameters and inputs) during the forward pass in order to be able to compute the gradients during the backward pass.

- How many additional variables would have to be stored if we backpropagated through 100 steps instead of 2?

- Now reusing the computation from the forward pass $y = \tanh(x)$ and $\frac{dy}{dx} = 1 - y^2$, mark the variables needed for backpropagation. How many additional variables would now have to be stored if we backpropagated through 100 steps instead of 2?

1.3 Exercise 3

Looking at your results from exercise 2, where does the dependency on the inputs x_1 and x_2 appear in $\frac{dz}{dW}$?

1.4 Exercise 4

The gradient of your loss function with respect to the the parameters of a recurrent neural network turns out to be large and the loss seems to not decrease. Which action can you take in order to continue training the same recurrent neural network? Name the technical term and shortly describe the precise effect of the measure form of a text or a formula!

1.5 Exercise 5

The following recurrent neural network is given:

$$h_t = \tanh(A_1 x_t + B_1 h_{t-1} + c_1)$$
$$k_t = \tanh(A_2 x_t + B_2 k_{t+1} + c_2)$$
$$y_t = h_t + k_t$$

- What type of recurrent neural network is presented?

- Name a task that this type of neural network would be suited for!

1.6 Exercise 6

The following recurrent neural network is given:

$$h_t = \tanh(a x_t + b h_{t-1} + c)$$
$$y_t = h_t$$
$$L = \sum_t y_t^2$$

There are two parameter sets:

(1) $a = 20$, $b = 1$, $c = 1$

(2) $a = 1$, $b = 1$, $c = 0$

The initial hidden state and input data are: $h_0 = 0$, $x_1 = 0.5$, $x_2 = -1$.

- Calculate the loss L over two time-steps for the given input data and the two parameter sets (1) and (2).

- Should you initialize your parameters with (1) or (2)? Explain based on your calculations!

(Round all results and intermediate results to two decimal places.)

1.7 Exercise 7

Before training your recurrent neural network, you decide to split long sequences into multiple shorter ones. Name one disadvantage and one advantage of your decision.

2 Coding Exercises

- Understand the provided Jupiter Notebook on classifying handwritten digits with a RNN.

3 Homework

- Answer the multiple choice question in Moodle (more than one answer options can be correct!).

- Write your own function to calculate the forward passes for the given recurrent neural networks in the Jupyter Notebook. Submit the outputs for the requested time-steps in Moodle.