Exercise 1 - RNN for Sentiment Analysis

Suppose we are training a vanilla RNN like below to determine whether a sentence expresses positive or negative sentiment. This RNN will be a character-level RNN where $x^{(1)}, \dots, x^{(T)}$ is the sequence of input characters. The RNN is given as follows:

$$h^{(t)} = \tanh\left(Ux^{(t)} + Wh^{(t-1)} + b\right)$$
$$y = \sigma\left(Vh^{(T)} + d\right)$$

(a) How many times do we need to apply the weight matrix U, W, and V?

(b) What are the shapes of the matrices U, W, and V?

(c) How many addition and multiplication operations are required to make a prediction? You can assume that no addition and multiplications are performed when applying the tanh and sigmoid activation functions.

Solution

(a) We will need to compute $h^{(t)}$ for $t = 1, \dots, T$. Each of this computation requires applying the weight matrices W and T once. The matrix V is only applied once at the end. Therefore, we need to apply W and U T times each and V once.

(b) The shape of U is $d_h \times d_x$, the shape of W is $d_h \times d_h$, and the shape of V is $d_y \times d_h$, where d_h is the dimensionality of the $h^{(i)}$ (i.e. $h^{(i)} \in \mathbb{R}^{d_h}$), d_x is the dimensionality of the inputs $x^{(i)}$, and d_y is the dimensionality of the ouput y.

(c) For each of the T steps, we need to perform two matrix-vector multiplications (one for $Ux^{(i)}$ and one for $Uh^{(i)}$) and two vector additions. To compute the output, we need one additional matrix-vector multiplication and one vector addition.

Exercise 2 - Scalar RNN

Suppose we have the following vanilla RNN network, where the inputs and hidden units are scalars.

$$h^{(t)} = \tanh\left(w \cdot h^{(t-1)} + u \cdot x^{(t-1)} + b_h\right)$$
$$y = \sigma\left(v \cdot h^{(T)} + b_y\right)$$

(a) Show that if $|w| < 1$, and the number of time steps T is large, then the gradient $\frac{\partial y}{\partial x^{(0)}}$ vanishes.

(b) Why is the result from Part (a) troubling?

Solution

(a) To make the sequence length T explicit in the notation, we will write y instead of y_T. Formally, what we have to show is

$$|w| < 1 \implies \lim_{T \to \infty} \frac{\partial y_T}{\partial x^{(0)}} = 0.$$