6.390 Introduction to Machine Learning

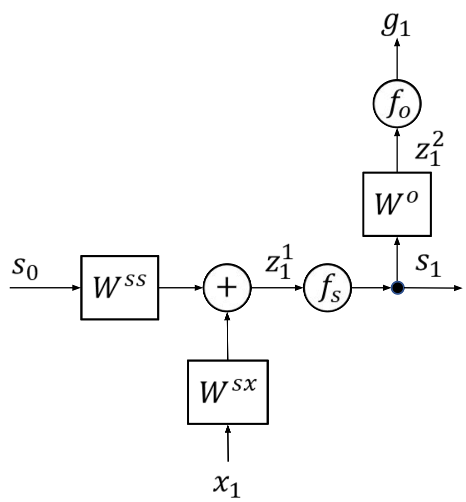
Recitation Week #8

Issued April 10, 2023

1. Consider the following recurrent neural network (RNN):

$$z_t^1 = W^{ss} s_{t-1} + W^{sx} x_t, \qquad s_t = f_s(z_t^1),$$
$$z_t^2 = W^o s_t, \qquad g_t = f_o(z_t^2),$$

where we have set biases to zero. Here, we have a data set $\mathcal{D}_n = \left\{ \left(x^{(i)}, y^{(i)}\right) \right\}_{i=1}^q$ consisting of input-output sequence pairs. The i^{th} sequence has length $n^{(i)}$. The output of the RNN is designated by a sequence $g^{(i)} = \text{RNN}(x^{(i)}; W)$, where W is an object which consists of the weight matrices $\{W^{ss}, W^{sx}, W^o\}$. In the figure below is a visualization of one stage of the unrolled RNN.



(a) Assume our first RNN, call it RNN-A, has s_t, x_t, g_t all being vectors. Let s_t be of shape 2×1, x_t of shape 3×1 and g_t of shape 4×1. In addition, the activation functions are $f_s(z) = z$ and $f_o(z) = z$.

i. For RNN-A, give dimensions of the following vectors and matrices:

W^{ss}: ____________ W^{sx}: ____________ W^o: ____________

z_t^1: ____________ z_t^2: ____________

ii. Consider a particular RNN-A where all the elements of W^{ss}, W^{sx}, and W^o are 1. Let's look at an input sequence x of length 1. The first (and only) element of this sequence is $x_1 = [1, 0, 0]^\top$. Our initial state is $s_0 = [1, 1]^\top$. Since the input sequence had length 1, the output sequence will also have length 1. What is the output of RNN-A, g_1?

(b) Using the structure of RNN-A, we wish to implement a couple of different functionalities.

We've got a lot of dimensions we're using. We have a sequence x of vectors (and we could have multiple sequences, which we would then distinguish with a superscript). The t^{th} element of our sequence is the vector x_t, which in turn has multiple elements. Here, we'll use bracket notation $x_t[i]$ to denote the i^{th} element of the vector x_t.

i. First, we want to implement weight matrices and initial states such that the output g_t will be:

$$g_t = \begin{bmatrix} x_t[1] \\ 0 \\ 0 \\ \sum_{j=1}^{t} \sum_{i=1}^{3} x_j[i] \end{bmatrix}.$$

That is, the first element of g_t is the first element of x_t, and the last element of g_t is the sum of all of the elements of x over all inputs x_j, for $j = 1, \ldots, t$.

Define s_0, W^{ss}, W^{sx}, and W^o necessary to implement the behavior described above.

ii. Now, we want weight matrices and initial state to implement:

$$g_t = \begin{bmatrix} x_t[1] \\ x_t[2] \\ x_t[3] \\ \sum_{j=1}^{t} \sum_{i=1}^{3} x_j[i] \end{bmatrix}.$$

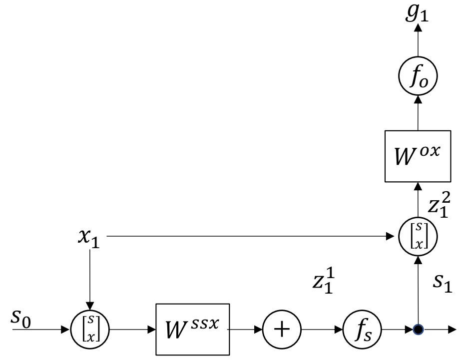
That is, to have the output $g_t[2] = x_t[2]$ and $g_t[3] = x_t[3]$, in addition to the outputs from part (b) i. If we keep the state at size 2×1 and the activation functions as $f_s(z) = z$ and $f_o(z) = z$, is it possible to implement this with the RNN-A structure? Why or why not?

(c) Now consider a modified RNN, call it RNN-B, that does the following:

$$z_t^1 = \begin{bmatrix} s_{t-1} \\ x_t \end{bmatrix}, \qquad s_t = f_s\left(W^{ssx} z_t^1\right),$$

$$z_t^2 = \begin{bmatrix} s_t \\ x_t \end{bmatrix}, \qquad g_t = f_o\left(W^{ox} z_t^2\right),$$

where s_t, x_t, g_t are all vectors. Let s_t be of shape 2×1, x_t of shape 3×1 and g_t of shape 4×1. Then, $\begin{bmatrix} s_{t-1} \\ x_t \end{bmatrix}$ and $\begin{bmatrix} s_t \\ x_t \end{bmatrix}$ are the concatenation of the two column vectors into a new column vector. In addition, the activation functions are $f_s(z) = z$ and $f_o(z) = z$. In the figure below is a visualization of one stage of this RNN-B.



i. For RNN-B, give dimensions of the following matrices:

W^{ssx}: ____________ W^{ox}: ____________

z_t^1: ____________ z_t^2: ____________

ii. Now, would it be possible for RNN-B to implement the functionality discussed in part (b) ii? If so, define s_0, W^{ssx}, and W^{ox} necessary to implement the behavior.

iii. Instead of using RNN-B, could we change the state space representation s_t and weights in our standard RNN structure (RNN-A) to achieve the capabilities of RNN-B?