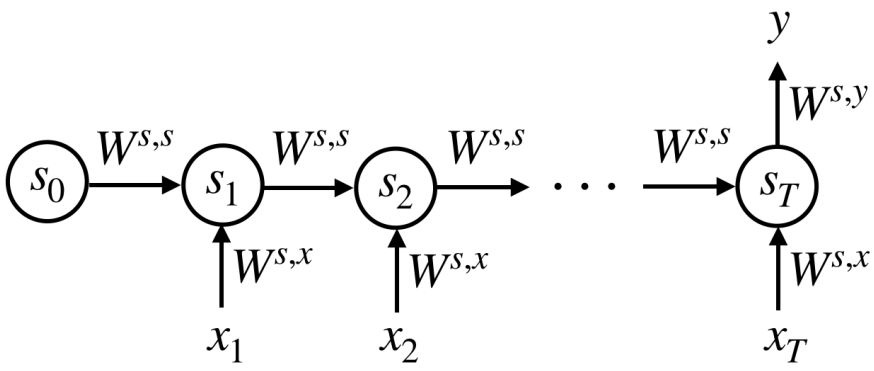
Problem 6



Recurrent neural networks (RNN) can be used as classification models for time series data. Here we have a simple RNN as shown in the figure above, where

$$s_t = f_1\left(W^{s,s}s_{t-1} + W^{s,x}x_t\right), \quad t = 1, 2, \ldots, T$$

and

$$y = f_2\left(W^{s,y}s_T + W_0\right)$$

We assume all offsets are 0 except W_0 for the final output layer and we decide the two activation functions to be:

$$f_1(z) = \mathrm{RELU}(z) = max(0, z)$$

and

$$f_2(z) = \mathrm{sign}(z) = \begin{cases} 1, & if \quad z \geq 0 \\ 0, & if \quad z < 0 \end{cases}$$

Note that the $\mathrm{RELU}(z)$ can be applied elementwise if z is a vector.

Suppose we want to apply this model to classify sentences into different categories (e.g. positive/negative sentiment), we need to encode each word in a sentence into a vector as the input x_t to the model. One way to do this is to represent the tth word as a column vector of length $|V|$, where V is the set of the entire vocabulary. The ith element of x_t is 1 if the word is the ith word in the vocabulary and all other elements are zero.

6. (1)

2.0/2 points (graded)

We first explore a simple scenario where our vocabulary contains only 2 words, $V = \{A, B\}$. Let $s_t \in \mathbb{R}^2$ and we set the initial state s_0 and the weights before the last layer as follows:

$$s_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad W^{s,s} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad W^{s,x} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Now given 3 training sentences: AA, ABB, BAA

Encode each of them into a sequence of vectors. As an example, the sentence AA is encoded as $x^{(1)} = (x_1^{(1)}, x_2^{(1)})$, where $x_1^{(1)} = x_2^{(1)} = [1, 0]^T$.

(To enter the sequence above, type `[[1,0],[1,0]]`.)

Now encode the other 2 sentences into $x^{(2)}$ and $x^{(3)}$.

$x^{(2)} =$ [[1, 0], [0, 1], [0, 1]] ✔ **Answer:** [[1,0],[0,1],[0,1]]

$x^{(3)} =$ [[0, 1], [1, 0], [1, 0]] ✔ **Answer:** [[0,1],[1,0],[1,0]]

Solution:

A is encoded as $[1,0]^T$ and B is encoded as $[0,1]^T$, so we have

$$x^{(2)} = (x_1^{(2)}, x_2^{(2)}, x_3^{(2)}) = ([1,0]^T, [0,1]^T, [0,1]^T)$$
$$x^{(3)} = (x_1^{(3)}, x_2^{(3)}, x_3^{(3)}) = ([0,1]^T, [1,0]^T, [1,0]^T)$$

Submit You have used 1 of 5 attempts

ℹ Answers are displayed within the problem

6. (2)

3.0/3 points (graded)

Now compute the final hidden state $s_T^{(1)}, s_T^{(2)}, s_T^{(3)}$ for each of the three senteces AA, ABB, BAA in this RNN.

(Enter `[0,0]` for $S_T = [0,0]^T$.)

$s_T^{(1)} =$ [0, 0] ✔ **Answer:** [0,0]

$s_T^{(2)} =$ [0, 2] ✔ **Answer:** [0,2]

$s_T^{(3)} =$ [0, 1] ✔ **Answer:** [0,1]

Solution:

Using $s_t = RELU\left(W^{s,s} s_{t-1} + W^{s,x} x_t\right)$, we can compute

$$s_T^{(1)} = [0,0]^T$$
$$s_T^{(2)} = [0,2]^T$$
$$s_T^{(3)} = [0,1]^T$$

Submit You have used 1 of 5 attempts

ℹ Answers are displayed within the problem

6. (3)

1/1 point (graded)

Fixing s_0, $W^{s,s}$, $W^{s,x}$ and by only learning the linear classifier in the final layer, can this RNN separate the 3 examples regardless of how they were labeled?

- Yes
- ◉ No

✔

Solution:

No. As the final layer is a linear classifier and $s_T^{(1)}$, $s_T^{(2)}$, $s_T^{(2)}$ are collinear points, they are not linear seperable in general. A concrete example is when $y^{(1)} = y^{(2)} = 1$ and $y^{(3)} = -1$.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

6. (4)

1/1 point (graded)
A simpler model to classify sentences is to represent the entire sentence into a vector z and apply a linear model on z, i.e.

$$y = sign\left(W^{z,y} z\right)$$

The vector z has length $|V|$ and the ith element of z is the count of how many times the ith word appears in the sentence. For example, the sentence ABA with $V = \{A, B\}$ will be encoded as $z = [2, 1]^T$. If we want the RNN we described earlier to match the output of this linear model given any input sentences, Which of the following is a possible setting of the weights and initial state s_0 of the RNN? Check all that apply.

Here $c^{|V|}$ stands for a vector of length $|V|$ in which every element is c (e.g. $[1, 1, 1]^T$ if $c = 1$ and $|V| = 3$), $I_{|V|}$ stands for the identity matrix of size $|V|$.

- [ ] $s_0 = 1^{|V|}, W^{s,s} = I_{|V|}, W^{s,x} = I_{|V|}, W^{s,y} = -W^{z,y}, W_0 = \sum_i W_i^{z,y}$
- [x] $s_0 = 1^{|V|}, W^{s,s} = I_{|V|}, W^{s,x} = I_{|V|}, W^{s,y} = W^{z,y}, W_0 = -\sum_i W_i^{z,y}$
- [ ] $s_0 = 0^{|V|}, W^{s,s} = I_{|V|}, W^{s,x} = -I_{|V|}, W^{s,y} = -W^{z,y}, W_0 = 0$
- [x] $s_0 = 0^{|V|}, W^{s,s} = I_{|V|}, W^{s,x} = I_{|V|}, W^{s,y} = W^{z,y}, W_0 = 0$

Solution:

By setting $W^{s,s} = I_{|V|}$ and $W^{s,x} = I_{|V|}$, we have $s_t = RELU\left(s_{t-1} + x_t\right)$.
If we initialize $s_0 = 0^{|V|}$, then s_T will be the same as z,
thus $W^{s,y} = W^{z,y}$ and $W_0 = 0$.
If we initialize $s_0 = 1^{|V|}$,
then $s_T = z + 1^{|V|}$.
To make $W^{z,y} z = W^{s,y}\left(z + 1^{|V|}\right) + W_0$,
we have $W^{s,y} = W^{z,y}$ and $W_0 = -W^{s,y} \cdot 1^{|V|} = -\sum_i W_i^{z,y}$

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

6. (5)

0/1 point (graded)
Now suppose we want to use indicators instead of counts for the vector z. That is the ith element of z will be 1 if the ith word appears anywhere in the sentence. Which of the following is a possible setting of the weights and initial state s_0 of the RNN? Check all that apply.

Note (Sept 8): In the choices below, W_0 is a scalar, and the summation $\sum_i W_i^{z,y}$ over i is summing over the elements of the vector $W^{z,y}$.

- [ ] $s_0 = 1^{|V|}, W^{s,s} = I_{|V|}, W^{s,x} = -I_{|V|}, W^{s,y} = W^{z,y}, W_0 = -\sum_i W_i^{z,y}$
- [ ] $s_0 = 1^{|V|}, W^{s,s} = I_{|V|}, W^{s,x} = -I_{|V|}, W^{s,y} = -W^{z,y}, W_0 = \sum_i W_i^{z,y}$ ✔
- [ ] $s_0 = 0^{|V|}, W^{s,s} = I_{|V|}, W^{s,x} = -I_{|V|}, W^{s,y} = -W^{z,y}, W_0 = 0$
- [x] $s_0 = 0^{|V|}, W^{s,s} = I_{|V|}, W^{s,x} = I_{|V|}, W^{s,y} = W^{z,y}, W_0 = 0$

✖ Option 4 in 6.(5) will give the same output as z encoded in 6.(4), but clearly not the z as described in 6.(5).

Solution:

As the RELU activation function will map all non-positive input to 0, so whenever a word appears, we can minus the corresponding state by 1. If we set the initial state to be 1, representing a word does not appear, then as long as a word appears, no matter how many times, the state will be 0.

With this idea, we choose $s_0 = 1^{|V|}, W^{s,s} = I_{|V|}$ and $W^{s,x} = -I_{|V|}$. By doing so, we have $s_T = 1^{|V|} - z$.

To make $W^{z,y}z = W^{s,y}\left(1^{|V|} - z\right) + W_0$,

We have $W^{s,y} = -W^{z,y}$ and $W_0 = \sum_i W_i^{z,y}$.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Error and Bug Reports/Technical Issues

Topic: Final exam (1 week):Final Exam / Problem 6

Basically for Option 4, f_1 would return a vector indicating the number of occurrences of A and B. When $W^{s,y} = W^{z,y}$, it's not difficult to tell that the results before applying the sign function are not the same.

Yes the final output (the sign function output) might be the same if the learned values of all the elements of $W^{z,y}$ are positive. But it's not guaranteed.

The correct answer Option 2 is different. The calculated S_t is:

$f_1 = 0$ if the correspondent symbol has at least one occurrence.

$f_1 = 1$ if the correspondent symbol has no occurrence.

So we actually have $f_1 = 1 - z$ (Note: 1 is a vector of 1s). Given the condition in Option 2 that $W_0 = \sum_i W_i^{z,y}$, we have:

$W^{s,y} * S_t + W_0 = -W^{z,y} * (1 - z) + W_0 = W^{z,y} * z - \sum_i W_i^{z,y} + W_0 = W^{z,y} * z$

This would hold no matter what the values of the elements of $W^{z,y}$ are.

posted a day ago by sean_s_wang

Add a comment

Winston_Dai (Staff)
a day ago

Option 4 in 6.(5) will give the same output as z encoded in 6.(4), but clearly not the z as described in 6.(5).