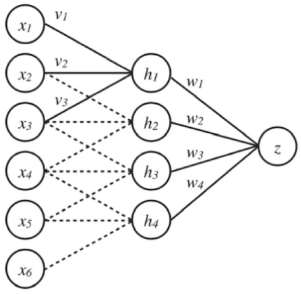
Assignment 4

Benny Chen

November 30, 2023

1 Simple Backward Propagation

Consider this convolutional neural network architecture.



In the first layer, we have a one-dimensional convolution with a single filter of size 3 such that

$$h_i = s(\sum_{j=1}^{3} v_j x_{i+j-1}) \tag{1}$$

The second layer is fully connected, such that

$$z = \sum_{i=1}^{4} w_i h_i \tag{2}$$

The hidden units' activation function (x) is the logistic (sigmoid) function with derivative

$$s'(x) = s(x)(1 - s(x)) \tag{3}$$

The output unit is linear (no activation function).
We perform gradient descent on the loss function

$$R = (y - z)^2 \tag{4}$$

where y is the training label for x.

(a) What is the total number of parameters in this neural network? Recall that convolutional layers share weights. There are no bias terms.

(b) Compute $\partial R/\partial w_i$. (Hint: in terms of y, z, and h_i)

(c) Compute $\partial R/\partial v_j$. (Hint: in terms of w_i, h_i, and x_{i+j-1})

Answer:

(a) There are 3 parameters in the first layer and 4 parameters in the second layer. Therefore, there are 7 parameters in total.

(b)
$$\frac{\partial R}{\partial w_i} = -2(y-z)h_i$$

(c)
$$\frac{\partial R}{\partial v_j} = \frac{\partial R}{\partial z}\frac{\partial z}{\partial h_i}\frac{\partial h_i}{\partial v_j} = -2(y-z)\sum_{i=1}^{4} w_i h_i(1-h_i)x_{i+j-1}$$

2 Convolution Neural Network

Consider a convolutional neural network (CNN) for reading the handwritten MNIST letters, which are 28×28 timages. Suppose the first hidden layer is a convolutional layer with 20 different 5×5 filters, applied to the input image with a stride of 1. Each filter has a bias weight. No padding operation is applied.

(a) How many weights (parameters) does this CNN layer use?

(b) What will be the output dimension of feature map after this layer?

Answer:

(a)
$$\text{Number of weights} = 20 \times (5 \times 5 + 1) = 520$$

(b)
$$\text{Output dimension} = 28 - 5 + 1 = 24$$

3 LSTM True or False plus Explanation

Suppose that here are the defining equations for a LSTM cell (might differ from the lecture note).

$$i_t = \sigma(W^{(i)}x_t + U^{(i)}h_{t-1}) \tag{5}$$

$$f_t = \sigma(W^{(f)}x_t + U^{(f)}h_{t-1}) \tag{6}$$

$$o_t = \sigma(W^{(o)}x_t + U^{(o)}h_{t-1}) \tag{7}$$

$$\tilde{c}_t = \tanh(W^{(c)}x_t + U^{(c)}h_{t-1}) \tag{8}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \tag{9}$$

$$h_t = o_t \circ \tanh(c_t) \tag{10}$$

Here the symbol $\circ$ denotes element-wise multiplication and that $\sigma(\cdot)$ denotes the sigmoid function.
Provide True or False for the following and briefly justify your answer

(a) If x_t is the 0 vector, then $h_t = h_{t-1}$.

(b) The entries of f_t, i_t, o_t are non-negative.

Answer:

(a) False. If x_t is the 0 vector, then $i_t = f_t = o_t = 0$. Therefore, $c_t = c_{t-1}$ and $h_t = 0$

(b) True. The sigmoid function has a range of $[0, 1]$.