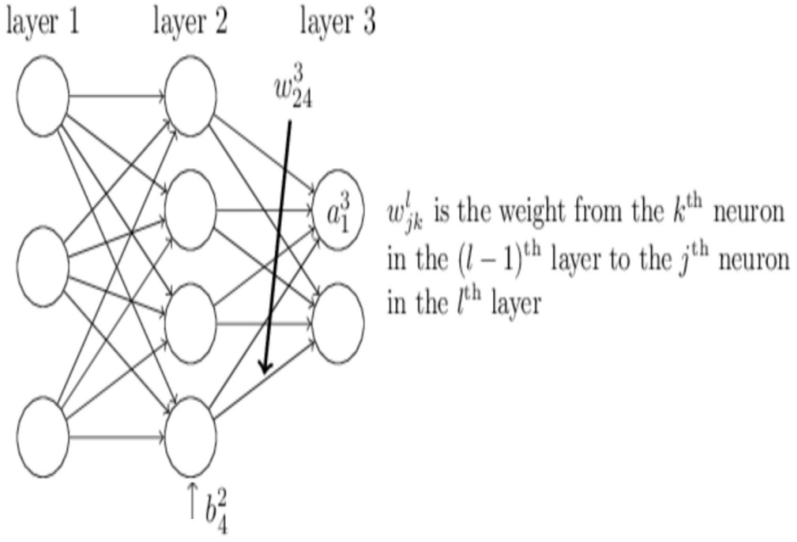
3. Backpropagation

Homework due Mar 29, 2023 08:59 -03 Completed

One of the key steps for training multi-layer neural networks is stochastic gradient desc
back-propagation algorithm to compute the gradient of the loss function with respect t

Consider the L-layer neural network below:



In the following problems, we will the following notation: b^l_j is the bias of the j^{th} neuron i
activation of j^{th} neuron in the l^{th} layer, and w^l_{jk} is the weight for the connection from the
layer to the j^{th} neuron in the l^{th} layer.

If the activation function is f and the loss function we are minimizing is C, then the equa
network are:

$$a^l_j = f\left(\sum_k w^l_{jk} a^{l-1}_k + b^l_j\right)$$

$$\text{Loss} = C(a^L)$$

Note that notations without subscript denote the corresponding vector or matrix, so th
of the l^{th} layer, and w^l is the weights matrix in l^{th} layer.

For $l = 1, \ldots, L$.

Computing the Error

1/2 points (graded)

Let the weighted inputs to the d neurons in layer l be defined as $z^l \equiv w^l a^{l-1} + b^l$, where z

- $f'(z_j^L)$

✔

What is δ_j^l for all $l \neq L$?

- $\sum_k w_{kj}^{l+1} \delta_k^{l+1} f'(z_j^l)$
- $\delta_k^{l+1} f'(z_j^l)$
- $\sum_k w_{jk}^{l-1} \delta_j^{l-1} f'(z_j^l)$
- $\sum_k w_{kj}^{l+1} \delta_k^{l+1} f(z_j^l)$

✖

Submit You have used 2 of 2 attempts

Parameter Derivatives

2/2 points (graded)

During SGD we are interested in relating the errors computed by back-propagation to t

interest: the partial derivatives of the loss with respect to our parameters. Here that is

What is $\frac{\partial C}{\partial w_{jk}^l}$? Write in terms of the variables a_k^{l-1}, w_j^l, b_j^l, and δ_j^l if necessary.

Example of writing superscripts and subscripts:

`delta_j^l` for δ_j^l

`w_(jk)^l` for w_{jk}^l

$\frac{\partial C}{\partial w_{jk}^l} =$ a_k^(l-1) * delta_j^l ✔

$\frac{\partial C}{\partial}$?

What is the derivative of the sigmoid function, $\sigma(z) = \frac{1}{1+e^{-z}}$? Please write your answer in

1 / (1 + e^(-z)) * (1 - 1 / (1 + e^(-z))

Which of the following is true of $\sigma'(z)$ as $||z||$ gets large?

- Its magnitude becomes large.
- Its magnitude becomes small.
- It suffers from high variance.

What is the derivative of the ReLU function, $\text{ReLU}(z) = \max(0, z)$ for $z > 0$?

1

For $z < 0$?

0

? STANDARD NOTATION

Submit You have used 1 of 5 attempts

Simple Network

4/4 points (graded)

Consider a simple 2-layer neural network with a single neuron in each layer. The loss fu

loss: $C = \frac{1}{2}(y - t)^2$, where y is the prediction and t is the target.

Starting with input x we have:

What are the derivatives with respect to the parameters?

$\frac{\partial C}{\partial w_1} =$ 2.081 ✔

$\frac{\partial C}{\partial w_2} =$ -0.004 ✔

$\frac{\partial C}{\partial b} =$ -0.139 ✔

? STANDARD NOTATION

Submit You have used 1 of 5 attempts

SGD

1/1 point (graded)

Referring to the previous problem, what is the update rule for w_1 in the SGD algorithm w terms of w_1, η, and $\frac{\partial C}{\partial w_1}$; enter the latter as `(partialC)/(partialw_1)`, noting the lack of spa names:

Next $w_1 =$ w_1 - eta*(partialC)/(partialw_1) ✔

? STANDARD NOTATION

Submit You have used 1 of 5 attempts

Discussion

Topic: Unit 3. Neural networks (2.5 weeks):Homework 3 / 3. Backpropagation

‹ Previous Next ›

Show all posts

Computing the error - question on solution

? Hi team. Any hint to answer the “Computing the Error” part “b” question?

I don’t understand what the sentence "What is delta_j^l for all 'l' different than 'L'" meant? Thanks in advance

SGD: enter step size as 'eta'

in case anyone runs into the same issue. I keep entering '\eta' which is rendered correctly, but not parsed.

? Is C a vector or scalar?

It might be a silly question but I am very confused. I believe that the loss C is a vector because it is f(a^L = ve

? Simple Network w_2 and bias derivatives

I got the first derivative correct but getting the next 2 wrong which is very odd because w_2 and the bias are

? Simple Network, invalid syntax

What are the variables to use to get the derivatives? for ex I got dC/dw_1 = [removed by staff] How to write c