2. Back-propagation Algorithm

Bookmark this page

Exercises due Oct 28, 2020 05:29 IST *Completed*

Back-propagation Algorithm

Start of transcript. Skip to the end.

0:00 / 0:00 1.25x

Welcome back.

Today, we're going to be talking about feed-forward neural

networks and in particular how to learn them from data.

If you recall, feed-forward neural networks

with multiple hidden layers mediating the calculation

Video

Download video file

Transcripts

Download SubRip (.srt) file

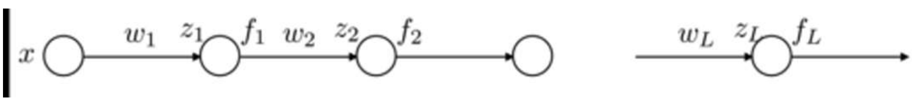
Download Text (.txt) file

Once we set up the architecture of our (feedforward) neural network, our goal will be to find weight parameters that minimize our loss function. We will use the **stochastic gradient descent algorithm** (which you learned in Lecture 4 and revisited in lecture 5) to carry out the optimization.

This involves computing the gradient of the loss function with respect to the weight parameters.

Since the loss function is a long chain of compositions of activation functions with the weight parameters entering at different stages, we will break down the computation of the gradient into different pieces via the chain rule; this way of computing the gradient is called the back-propagation algorithm.

In the following problems, we will explore the main step in the stochastic gradient descent algorithm for training the following simple neural network from the video:



This simple neural network is made up of L hidden layers, but each layer consists of only one unit, and each unit has activation function f.
As usual, x is the input, z_i is the weighted combination of the inputs to the i^{th} hidden layer. In this one-dimensional case, weighted combination reduces to products:

$$\begin{aligned} z_1 &= x w_1 \\ \text{for } i = 2 \ldots L: \quad z_i &= f_{i-1} w_i \quad \text{where } f_{i-1} = f(z_{i-1}) . \end{aligned}$$

We will use the following loss function:

$$\mathcal{L}(y, f_L) = (y - f_L)^2$$

where y is the true value, and f_L is the output of the neural network.

Gradient Descent Update

1/1 point (graded)

Let η be the learning rate for the stochastic gradient descent algorithm.

Recall that our goal is to tune the parameters of the neural network so as to minimize the loss function. Which of the following is the appropriate update rule for the paramter w_1 in the stochastic gradient descent algorithm?

- [x] $w_1 \leftarrow w_1 - \eta \cdot \nabla_{w_1} \mathcal{L}(y, f_L)$
- [ ] $w_1 \leftarrow w_1 + \eta \cdot \nabla_{w_1} \mathcal{L}(y, f_L)$
- [ ] $w_1 \leftarrow \eta \cdot \nabla_{w_1} \mathcal{L}(y, f_L)$
- [ ] $w_1 \leftarrow -\eta \cdot \nabla_{w_1} \mathcal{L}(y, f_L)$

✔

Solution:

The value of a function is non-decreasing in the direction of its gradient from any given point in its domain. Since our goal is to tune the parameters of the neural network so as to minimize the loss, we update the weights in the direction opposite to that of the gradient from any point.

The learning rate η controls the magnitude of the update made during gradient descent.

The final update for any parameter θ with gradient $\nabla_\theta \mathcal{L}$ would be given as follows:

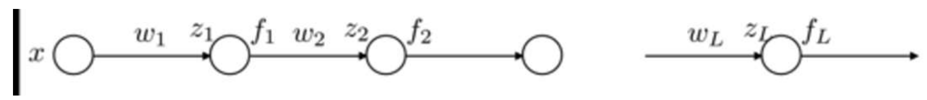
$$\theta \leftarrow \theta - \eta \cdot \nabla_\theta \mathcal{L}.$$

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Recursive Expression - Part I

1/1 point (graded)



As above, let $\mathcal{L}(y, f_L)$ denote the loss function as a function of the predictions f_L and the true label y. Recall

$$\begin{aligned} z_1 &= x w_1 \\ \text{for } i = 2 \dots L: \quad z_i &= f_{i-1} w_i \quad \text{where } f_{i-1} = f(z_{i-1}). \end{aligned}$$

Let $\delta_i = \dfrac{\partial \mathcal{L}}{\partial z_i}$.

The first step to updating any weight w is to calculate $\dfrac{\partial \mathcal{L}}{\partial w}$.

Which of the following option(s) is/are correct expression(s) for $\dfrac{\partial \mathcal{L}}{\partial w_1}$?

(Choose all that apply.)

- [ ] $\frac{\partial \mathcal{L}}{\partial w_1} = \frac{\partial w_1}{\partial z_1} \cdot \frac{\partial \mathcal{L}}{\partial z_1}$
- [x] $\frac{\partial \mathcal{L}}{\partial w_1} = \frac{\partial z_1}{\partial w_1} \cdot \frac{\partial \mathcal{L}}{\partial z_1}$
- [x] $\frac{\partial \mathcal{L}}{\partial w_1} = x \cdot \delta_1$
- [ ] $\frac{\partial \mathcal{L}}{\partial w_1} = x + \delta_1$

✔

Solution:

From the chain rule we have:

$$\frac{\partial \mathcal{L}}{\partial w_1} = \frac{\partial z_1}{\partial w_1} \frac{\partial \mathcal{L}}{\partial z_1}.$$

Since $z_1 = w_1 x,$ we get

$$\frac{\partial z_1}{\partial w_1} = \frac{\partial (w_1 . x)}{\partial w_1} = x.$$

Therefore, equivalently,

$$\frac{\partial \mathcal{L}}{\partial w_1} = x\delta_1.$$

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Recursive Expression - Part II

1/1 point (graded)
As above, let $\mathcal{L}(y, f_L)$ denote the loss function as a function of the predictions f_L and the true label y. Let $\delta_i = \frac{\partial \mathcal{L}}{\partial z_i}$.

In this problem, we derive a recurrence relation between δ_i and δ_{i+1}

Assume that f is the hyperbolic tangent function:

$$\begin{aligned} f(x) &= \tanh(x) \\ f'(x) &= (1 - \tanh^2(x)). \end{aligned}$$

Which of the following option is the correct expression for δ_1 in terms of δ_2?

- (●) $\delta_1 = (1 - f_1^2) \cdot w_2 \cdot \delta_2$
- ( ) $\delta_1 = (1 - f_1^2) \cdot w_1 \cdot \delta_2$

$\bigcirc$ $\delta_1 = (1 - f_2^2) \cdot w_2 \cdot \delta_2$

$\bigcirc$ $\delta_2 = (1 - f_1^2) \cdot w_2 \cdot \delta_1$

✔

Submit You have used 1 of 2 attempts

Final Expression of the Gradient

1/1 point (graded)

As above, let $\mathcal{L}(y, f_L)$ denote the loss function as a function of the predictions f_L and the true label y. Let $\delta_i = \frac{\partial \mathcal{L}}{\partial z_i}$.

In this problem, we unroll the recurrence expression for $\frac{\partial \mathcal{L}}{\partial w_1}$. We will use the loss function

$$\mathcal{L}(y, f_L) = (y - f_L)^2.$$

Compute $\frac{\partial \mathcal{L}}{\partial w_1}$ and select the correct option from below.

$\bigcirc$ $\frac{\partial \mathcal{L}}{\partial w_1} = x(1 - f_2^2) \cdots (1 - f_L^2) w_2 w_3 \cdots w_L . (2(f_L - y))$

$\bigcirc$ $\frac{\partial \mathcal{L}}{\partial w_1} = x(1 - f_1^2)(1 - f_2^2) \cdots (1 - f_L^2) w_1 w_2 w_3 \cdots w_L (2(f_L - y))$

$\circledcirc$ $\frac{\partial \mathcal{L}}{\partial w_1} = x(1 - f_1^2)(1 - f_2^2) \cdots (1 - f_L^2) w_2 w_3 \cdots w_L (2(f_L - y))$

$\bigcirc$ $\frac{\partial \mathcal{L}}{\partial w_1} = x w_2 w_3 \cdots w_L (2(f_L - y))$

✔

Solution:

From the previous problem, we know the following:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial w_1} &= x \cdot \delta_1 \\ \delta_1 &= (1 - f_1^2) \cdot w_2 \cdot \delta_2. \end{aligned}$$

Similarly, $\delta_2, \delta_3 \ldots \delta_L$ can be given as follows:

$$\begin{aligned} \delta_2 &= (1 - f_2^2) \cdot w_3 \cdot \delta_3 \\ \delta_3 &= (1 - f_3^2) \cdot w_4 \cdot \delta_4 \\ &\vdots \\ \delta_{L-1} &= (1 - f_{L-1}^2) \cdot w_L \cdot \delta_L \\ \delta_L &= \frac{\partial \mathcal{L}}{\partial z_L} \\ \delta_L &= \frac{\partial \mathcal{L}}{\partial f_L} \cdot \frac{\partial f_L}{\partial z_L} \\ \delta_L &= \frac{\partial (f_L - y)^2}{\partial f_L} \frac{\partial f_L}{\partial z_L} \\ \delta_L &= 2(f_L - y) \frac{\partial f_L}{\partial z_L} \end{aligned}$$

$$\delta_L = 2\left(f_L - y\right)\left(1 - f_L^2\right).$$

Plugging the above equations into the expression for $\frac{\partial \mathcal{L}}{\partial w_1}$ we get:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial w_1} &= x \cdot \delta_1 \\
\frac{\partial \mathcal{L}}{\partial w_1} &= x \cdot \left(1 - f_1^2\right) \cdot w_2 \cdot \delta_2 \\
\frac{\partial \mathcal{L}}{\partial w_1} &= x \cdot \left(1 - f_1^2\right) \cdot w_2 \cdot \left(1 - f_2^2\right) \cdot w_3 \cdot \delta_3 \\
&\vdots \\
\frac{\partial \mathcal{L}}{\partial w_1} &= x\left(1 - f_1^2\right)\left(1 - f_2^2\right)\cdots\left(1 - f_L^2\right) w_2 w_3 \cdots w_L \left(2\left(f_L - y\right)\right).
\end{aligned}$$

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 3 Neural networks (2.5 weeks):Lecture 9. Feedforward Neural Networks, Back Propagation, and Stochastic Gradient Descent (SGD) / 2. Back-propagation Algorithm

Add a Post

Show all posts by recent activity

- Video on Backpropagation by Statquest 3
  As always, a simpler view on the subject by StatQuest: https://www.youtube.com/watch?v=IN2XmBhILt4&t=143s
- I looked at the answer for " Final Expression of the Gradient " : the case of my "orphaned" negative sign. 9
  Community TA
- partial derivative with respect to a function 2
  Can anybody give an explanation on what does it even mean to take partial derivative with respect to a function like f1? Thanks.
- Interpretation of the solutions to the problems 3
  So I did the calculus, the algebra, the recursion, I got them right, but what do these solution mean? Is there some verbal interpretation t...
- Handwritten lecture notes 8
  Check out my notes for this lecture : https://drive.google.com/drive/folders/172YN9JMYWjb-6k6Sd3USa4TInUKSROWr?usp=sharing
- Notation for "Recursive Expression - Part II" 5
- Chain Rule - Chain rule and tree diagrams of multivariable functions 1
  Found a video about chain rule. https://www.youtube.com/watch?v=kCr13iTRN7E and https://www.youtube.com/watch?v=NO3AqAaAE6o
- Thank you for the notes after video segments 1
  whoever wrote them did me a huge favor
- Technically, aren't there L-1 hidden layers? 2
- Just a refresher of chain rule if needed (link) 2
  Hope it helps. https://www.youtube.com/watch?v=wl1myxrtQHQ
  Community TA
- [STAFF] Recursive Expression - Part II 3
  I might be mistaken, but can the staff check the grader for this question? I'm not sure why it's incorrect since I used that basis to get the ...
- Links to Lecture 4 and Lecture 5 are broken 4
  The links on this page are to an old version of the course ... could they be updated?

Previous

Next

edX

About
Affiliates
edX for Business
Open edX
Careers
News

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect

Blog
Contact Us
Help Center
Media Kit
Donate

GET IT ON Google play
Download on the App Store