Deep Learning (CS 637) Homework 1

Mitchell Dodson

February 28, 2024

1: Consider the model $y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2$. Calculate the gradient with respect to $\vec{w}$, and write down a gradient descent update rule for $\vec{w}$ with respect to $\boldsymbol{\nabla}_{\vec{w}} y$.

The gradient of y with respect to weight vector $\vec{w} :=< w_1, w_2, w_3, w_4 >$ is written...

$$\boldsymbol{\nabla}_{\vec{w}} y = \left\langle \frac{\partial y}{\partial w_1}, \frac{\partial y}{\partial w_2}, \frac{\partial y}{\partial w_3}, \frac{\partial y}{\partial w_4} \right\rangle \tag{1}$$
$$= \left\langle 1,\ x_1,\ x_2,\ x_1^2 \right\rangle$$

In general, a gradient update step for parameter w is written $w' = w - \eta \frac{\partial \text{Loss}}{\partial w}$ so that the magnitude of the parameter's change is directionally opposite and proportional to the loss gradient with respect to it. In this case, the weight update in terms of the model's gradient evaluated for a single sample is as follows...

$$\vec{w'} = \vec{w} - \eta \boldsymbol{\nabla}_{\vec{w}} y = \left\langle (w_0 - \eta),\ (w_1 - \eta x_1),\ (w_2 - \eta x_2),\ (w_3 - \eta x_1^2) \right\rangle \tag{2}$$

2: Implement backpropagation by hand for the simple network, showing all the intermediate steps and calculations towards the computation of the solution.

Terms

$$\sigma(x) := \frac{1}{1 + e^{-x}} \tag{3}$$

$$\text{Div}(\hat{y}, y) := |\hat{y} - y|^2 \tag{4}$$

$$y = 0.5 \tag{5}$$

$$\vec{x} = \langle -0.7,\ 1.2,\ 1.1,\ -2 \rangle \tag{6}$$

$$\vec{w} = \langle -1.7,\ 0.1,\ -0.6,\ -1.8,\ -0.2,\ 0.5 \rangle \tag{7}$$

Forward Pass

$$
\begin{aligned}
h_1 &= \sigma(x_1 w_1 + x_2 w_2) \\
&= \sigma(-0.7 \cdot -1.7 + 1.2 \cdot 0.1) \\
&= \sigma(1.31) = 0.78751
\end{aligned}
\tag{8}
$$

$$
\begin{aligned}
h_2 &= \sigma(x_3 w_3 + x_4 w_4) \\
&= \sigma(1.1 \cdot -0.6 + -2 \cdot -1.8) \\
&= \sigma(2.94) = 0.9498
\end{aligned}
\tag{9}
$$

$$
\begin{aligned}
\hat{y} &= \sigma(h_1 w_5 + h_2 w_6) \\
&= \sigma(0.78751 \cdot -0.2 + 0.9498 \cdot 0.5) \\
&= \sigma(0.317) = 0.57869
\end{aligned}
\tag{10}
$$

$$
\begin{aligned}
\text{Div}(y, \hat{y}) &= |y - \hat{y}|^2 \\
&= |0.5 - 0.57869|^2 = 0.006192
\end{aligned}
\tag{11}
$$

Backward Pass

$$
\boldsymbol{\nabla}_{\hat{y}}\text{Div} = 2\,|\hat{y} - y| = 2\,|0.5 - 0.57869| = 0.15738
\tag{12}
$$

$$
\begin{aligned}
\boldsymbol{\nabla}_{s_3}\text{Div} &= \boldsymbol{\nabla}_{\hat{y}}\text{Div} \cdot \boldsymbol{\nabla}_{s_3}\hat{y} \\
&= \boldsymbol{\nabla}_{\hat{y}}\text{Div} \cdot \frac{\partial}{\partial s_3}\sigma(s_3) \\
&= \boldsymbol{\nabla}_{\hat{y}}\text{Div} \cdot \sigma(s_3)(1 - \sigma(s_3)) \\
&= 0.15738 \cdot 0.57869(1 - 0.57869) \\
&= 0.03837
\end{aligned}
\tag{13}
$$

$$
\begin{aligned}
\boldsymbol{\nabla}_{h_1}\text{Div} &= \boldsymbol{\nabla}_{s_3}\text{Div} \cdot \boldsymbol{\nabla}_{h_1}s_3 \\
&= \boldsymbol{\nabla}_{s_3}\text{Div} \cdot \frac{\partial}{\partial h_1}(h_1 w_5 + h_2 w_6) \\
&= \boldsymbol{\nabla}_{s_3}\text{Div} \cdot w_5 \\
&= 0.03837 \cdot -0.2 \\
&= -0.007674
\end{aligned}
\tag{14}
$$

$$\begin{aligned}
\boldsymbol{\nabla}_{s_1}\text{Div} &= \boldsymbol{\nabla}_{h_1}\text{Div} \cdot \boldsymbol{\nabla}_{s_1} h_1 \\
&= \boldsymbol{\nabla}_{h_1}\text{Div} \cdot \frac{\partial}{\partial s_1}\sigma(s_1) \\
&= \boldsymbol{\nabla}_{h_1}\text{Div} \cdot \sigma(s_1)(1 - \sigma(s_1)) \\
&= -0.007674 \cdot 0.78751(1 - 0.78751) \\
&= -0.001284
\end{aligned} \tag{15}$$

$$\begin{aligned}
\boldsymbol{\nabla}_{w_1}\text{Div} &= \boldsymbol{\nabla}_{s_1}\text{Div} \cdot \boldsymbol{\nabla}_{w_1} s_1 \\
&= \boldsymbol{\nabla}_{s_1}\text{Div} \cdot \frac{\partial}{\partial w_1}(x_1 w_1 + x_2 w_2) \\
&= \boldsymbol{\nabla}_{s_1}\text{Div} \cdot x_1 \\
&= -0.001284 \cdot -0.7 \\
&= 0.0008989
\end{aligned} \tag{16}$$

The partial derivative of the divergence function with respect to w_1 given the sample is $\frac{\partial \text{Div}}{\partial w_1} = 0.0008989$.

3:   Write a program implementing backpropagation on the neural network, and report the partial derivatives of the divergence with respect to all of the weights.

All of my code for this assignment is available on github at:
`https://github.com/Mitchell-D/cs637hw1`