The University of Yonsei

Faculty of Industrial Engineering

Tsoding MachineLearning In C

Lim Dohyun

November 18, 2023

1 Introduction

안녕하세요 한국말이 처음 입니다skddsds.

2 Gradient Descent

$$C'(w) = \lim_{\epsilon \to 0} \frac{C(w+\epsilon) - C(w)}{\epsilon} \tag{1}$$

2.1 Twice

sequence of derivating C(w) with respect to w.

$$\begin{align}
C(w) &= \frac{1}{n}\sum_{i=1}^{n}(x_i w - y_i)^2 \tag{2}\\
C'(w) &= \left(\frac{1}{n}\sum_{i=1}^{n}(x_i w - y_i)^2\right)' \tag{3}\\
&= \frac{1}{n}\left(\sum_{i=1}^{n}(x_i w - y_i)^2\right)' \tag{4}\\
&= \frac{2}{n}\sum_{i=1}^{n}(x_i w - y_i)(x_i)' \tag{5}
\end{align}$$

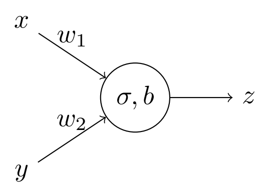
Cost funtction

$$C(w) = \frac{1}{n}\sum_{i=1}^{n}(x_i w - y_i)^2 \tag{6}$$

Derviative of Cost function

$$C'(w) = \frac{2}{n}\sum_{i=1}^{n}(x_i w - y_i)(x_i)' \tag{7}$$

2.2 One Neuron Model with 2 inputs



$$y = \sigma(xw_1 + yw_2 + b) \tag{8}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{9}$$

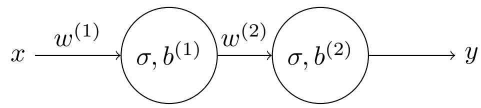
$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \tag{10}$$

2.2.1 Cost

subscript i is reffering to a sample number

$$
\begin{aligned}
a_i &= \sigma(x_i w_1 + y w_2 + b) && (11)\\
\partial_{w_1} a_i &= \partial_{w_1}(\sigma(x_i w_1 + y w_2 + b)) && (12)\\
&= a_i(1-a_i)\partial_{w_1}(x_i w_1 + y w_2 + b) && (13)\\
&= a_i(1-a_i)x_i && (14)\\
\partial_{w_2} a_i &= a_i(1-a_i)y_i && (15)\\
\partial_b a_i &= a_i(1-a_i) && (16)\\
C &= \frac{1}{n}\sum_{i=1}^{n}(a_i - y_i)^2 && (17)\\
\partial_{w_1} C &= \frac{1}{n}\sum_{i=1}^{n}\partial_{w_1}\left((a_i - z_i)^2\right) && (18)\\
&= \frac{1}{n}\sum_{i=1}^{n}2(a_i - z_i)\partial_{w_1} a_i && (19)\\
&= \frac{1}{n}\sum_{i=1}^{n}2(a_i - z_i)a_i(1-a_i)x_i && (20)\\
\partial_{w_2} C &= \frac{1}{n}\sum_{i=1}^{n}2(a_i - z_i)a_i(1-a_i)y_i && (21)\\
\partial_b C &= \frac{1}{n}\sum_{i=1}^{n}2(a_i - z_i)a_i(1-a_i) && (22)
\end{aligned}
$$

2.3 Two Neuron Model with 1 inputs



superscript i is reffering to a layer number

$$a_i^{(1)} = \sigma(x_i w^{(1)} + b^{(1)}) \tag{23}$$

$$\partial_{w^{(1)}} a_i^{(1)} = a_i^{(1)}(1 - a_i^{(1)}) x_i \tag{24}$$

$$\partial_{b^{(1)}} a_i^{(1)} = a_i^{(1)}(1 - a_i^{(1)}) \tag{25}$$

$$a_i^{(2)} = \sigma(a_i^{(1)} w^{(2)} + b^{(2)}) \tag{26}$$

$$\partial_{w^{(2)}} a_i^{(2)} = a_i^{(2)}(1 - a_i^{(2)}) a_i^{(1)} \tag{27}$$

$$\partial_{b^{(2)}} a_i^{(2)} = a_i^{(2)}(1 - a_i^{(2)}) \tag{28}$$

$$\partial_{a_i^{(1)}} a_i^{(2)} = a_i^{(2)}(1 - a_i^{(2)}) w^{(2)} \tag{29}$$

$$C^{(2)} = \frac{1}{n} \sum_{i=1}^{n} (a_i^{(2)} - y_i)^2 \tag{30}$$

$$\partial_{w^{(2)}} C^{(2)} = \frac{1}{n} \sum_{i=1}^{n} \partial_{w^{(2)}}((a_i^{(2)} - y_i)^2) \tag{31}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2(a_i^{(2)} - y_i) \partial_{w^{(2)}}(a_i^{(2)}) \tag{32}$$

$$= \frac{1}{n} \sum_{i=1}^{n} 2(a_i^{(2)} - y_i) a_i^{(2)}(1 - a_i^{(2)}) a_i^{(1)} \tag{33}$$

$$\partial_{b^{(2)}} C^{(2)} = \frac{1}{n} \sum_{i=1}^{n} 2(a_i^{(2)} - y_i) a_i^{(2)}(1 - a_i^{(2)}) \tag{34}$$

$$\partial_{a_i^{(1)}} C^{(2)} = \frac{1}{n} \sum_{i=1}^{n} 2(a_i^{(2)} - y_i) a_i^{(2)}(1 - a_i^{(2)}) w^{(2)} \tag{35}$$

$$e_i = a_i^{(1)} - \partial_{a_i^{(1)}} C^{(2)} \tag{36}$$

$$
\begin{align}
C^{(1)} &= \frac{1}{n}\sum_{i=1}^{n}(a_i^{(1)} - e_i)^2 \tag{37}\\
\partial_{w^{(1)}} C^{(1)} &= \partial_{w^{(1)}}\left(\frac{1}{n}\sum_{i=1}^{n}(a_i^{(1)} - e_i)^2\right) \tag{38}\\
&= \frac{1}{n}\sum_{i=1}^{n}\partial_{w^{(1)}}\left((a_i^{(1)} - e_i)^2\right) \tag{39}\\
&= \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(1)} - e_i)\partial_{w^{(1)}} a_i^{(1)} \tag{40}\\
&= \frac{1}{n}\sum_{i=1}^{n}2(a_i^{(1)} - (a_i^{(1)} - \partial_{a_i^{(1)}} C^{(2)}))\partial_{w^{(1)}} a_i^{(1)} \tag{41}\\
&= \frac{1}{n}\sum_{i=1}^{n}2(\partial_{a_i^{(1)}} C^{(2)})\partial_{w^{(1)}} a_i^{(1)} \tag{42}\\
&= \frac{1}{n}\sum_{i=1}^{n}2(\partial_{a_i^{(1)}} C^{(2)})a_i^{(1)}(1 - a_i^{(1)})x_i \tag{43}\\
\partial_{b^{(1)}} C^{(1)} &= \frac{1}{n}\sum_{i=1}^{n}2(\partial_{a_i^{(1)}} C^{(2)})\partial_{b^{(1)}} a_i^{(1)} \tag{44}\\
&= \frac{1}{n}\sum_{i=1}^{n}2(\partial_{a_i^{(1)}} C^{(2)})a_i^{(1)}(1 - a_i^{(1)}) \tag{45}
\end{align}
$$

2.4 Arbitrary Neurons Model with 1 inputs

Let's assume that we have m layers.

2.4.1 Feed-Forward

Let's assume that $a_i^{(0)}$ is x_i

$$
\begin{align}
a_i^{(l)} &= \sigma(a_i^{(l-1)} w^{(l)} + b^{(l)}) \tag{46}\\
\partial_{w^{(l)}} a_i^{(l)} &= a_i^{(l)}(1 - a_i^{(l)})a_i^{(l-1)} \tag{47}\\
\partial_{b^{(l)}} a_i^{(l)} &= a_i^{(l)}(1 - a_i^{(l)}) \tag{48}\\
\partial_{a_i^{(l-1)}} a_i^{(l)} &= a_i^{(l)}(1 - a_i^{(l)})w^{(l)} \tag{49}
\end{align}
$$

2.4.2 Back-Propagation - Tsoding's apporach

Let's denote $a_i^{(m)} - y_i$ as $\partial_{a_i^{(m)}} C^{(m+1)}$

$$C^{(l)} = \frac{1}{n}\sum_{i=1}^{n}(\partial_{a_i^{(l)}} C^{(l+1)})^2 \tag{50}$$

$$\partial_{w^{(l)}} C^{(l)} = \frac{1}{n}\sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)}) a_i^{(l)}(1 - a_i^{(l)}) a_i^{(l-1)} \tag{51}$$

$$\partial_{b^{(l)}} C^{(l)} = \frac{1}{n}\sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)}) a_i^{(l)}(1 - a_i^{(l)}) \tag{52}$$

$$\partial_{a_i^{(l-1)}} C^{(l)} = \frac{1}{n}\sum_{i=1}^{n} 2(\partial_{a_i^{(l)}} C^{(l+1)}) a_i^{(l)}(1 - a_i^{(l)}) w^{(l)} \tag{53}$$

2.4.3 Back-Propagation - Traditional apporach

above we denoted $a_i^{(m)} - y_i$ as $\partial_{a_i^{(m)}} C^{(m+1)}$ but here we are just going to use $a_i^{(m)} - y_i$

$$
\begin{aligned}
C^{(l)} &= (a_i^{(m)} - y)^2 && (54)\\
\partial_{a^{(m)}} C^{(l)} &= 2 \cdot (a_i^{(m)} - y) && (55)\\
\partial_{w^{(l)}} C^{(l)} &= \partial_{a_i^{(l)}} C^{(l+1)} \cdot a_i^{(l)}(1 - a_i^{(l)}) a^{(l-1)} && (56)\\
\partial_{b^{(l)}} C^{(l)} &= \partial_{a_i^{(l)}} C^{(l+1)} \cdot a_i^{(l)}(1 - a_i^{(l)}) && (57)\\
\partial_{a_i^{(l-1)}} C^{(l)} &= \partial_{a_i^{(l)}} C^{(l+1)} \cdot a_i^{(l)}(1 - a_i^{(l-1)}) w^{(l)} && (58)\\
& && (59)
\end{aligned}
$$