CENG 499 - Introduction to Machine Learning

Homework 1 - Part 1

Ahmet Eren Çolak
e2587921@ceng.metu.edu.tr

November 9, 2023

Derivatives for the Regression Problem

$$O_k^{(1)} = \sigma\left(\sum_{i=0} O_i^{(0)} \cdot a_{ik}^{(0)}\right) \tag{1}$$

$$O_0^{(2)} = \sum_{k=0} O_k^{(1)} \cdot a_{k0}^{(1)} \tag{2}$$

General rule for partial derivatives of weights in the second layer:

$$\frac{\partial E(x)}{\partial a_{k0}^{(1)}} = \frac{\partial (y - O_0^{(2)})^2}{\partial O_0^{(2)}} \frac{\partial O_0^{(2)}}{\partial a_{k0}^{(1)}} = 2(y - O_0^{(2)}) \frac{\partial O_0^{(2)}}{\partial a_{k0}^{(1)}}$$

Using the equation 2, it can be deduced that:

$$\frac{\partial O_0^{(2)}}{\partial a_{k0}^{(1)}} = O_k^{(1)}$$

$$\frac{\partial E(x)}{\partial a_{k0}^{(1)}} = 2(y - O_0^{(2)}) O_k^{(1)} \tag{3}$$

General rule for partial derivatives of weights in the first layer:

$$\frac{\partial E(x)}{\partial a_{ik}^{(0)}} = \frac{\partial (y - O_0^{(2)})^2}{\partial O_0^{(2)}} \frac{\partial O_0^{(2)}}{\partial O_k^{(1)}} \frac{\partial O_k^{(1)}}{\partial a_{ik}^{(0)}} = 2(y - O_0^{(2)}) a_{k0}^{(1)} \frac{\partial O_k^{(1)}}{\partial a_{ik}^{(0)}}$$

To find the derivative of the sigmoid function, it is written as:

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

$$\frac{d\sigma(x)}{dx} = (1 + e^{-x})^{-2} \cdot (e^{-x}) = \frac{1}{1 + e^{-x}} \frac{(1 + e^{-x}) - 1}{1 + e^{-x}} = \frac{1}{1 + e^{-x}} \left[\frac{1 + e^{-x}}{1 + e^{-x}} - \frac{1}{1 + e^{-x}}\right]$$

$$\frac{d\sigma(x)}{dx} = \sigma(x)(1 - \sigma(x))$$

Then general rule for partial derivatives of weights in the first layer can be found.

$$\frac{\partial E(x)}{\partial a_{ik}^{(0)}} = 2(y - O_0^{(2)}) a_{k0}^{(1)} \cdot \sigma(O_k^{(1)})(1 - \sigma(O_k^{(1)})) O_i^{(0)} \tag{4}$$

All update rules for weights in the second layer can be found using the equation 3.

$$a_{00}^{(1)} := a_{00}^{(1)} - \alpha \left[2(y - O_0^2) O_0^{(1)} \right]$$
$$a_{10}^{(1)} := a_{01}^{(1)} - \alpha \left[2(y - O_0^2) O_1^{(1)} \right]$$
$$a_{20}^{(1)} := a_{02}^{(1)} - \alpha \left[2(y - O_0^2) O_2^{(1)} \right]$$
$$a_{30}^{(1)} := a_{03}^{(1)} - \alpha \left[2(y - O_0^2) O_3^{(1)} \right]$$

All update rules for weights in the first layer can be found using the equation 4.

$$a_{01}^{(0)} := a_{01}^{(1)} - \alpha \left[2(y - O_0^2) a_{10}^{(1)} \sigma(O_1^{(1)})(1 - \sigma(O_1^{(1)})) O_0^{(0)} \right]$$
$$a_{02}^{(0)} := a_{02}^{(1)} - \alpha \left[2(y - O_0^2) a_{20}^{(1)} \sigma(O_2^{(1)})(1 - \sigma(O_2^{(1)})) O_0^{(0)} \right]$$
$$a_{03}^{(0)} := a_{03}^{(1)} - \alpha \left[2(y - O_0^2) a_{30}^{(1)} \sigma(O_3^{(1)})(1 - \sigma(O_3^{(1)})) O_0^{(0)} \right]$$

$$a_{11}^{(0)} := a_{11}^{(1)} - \alpha \left[2(y - O_0^2) a_{10}^{(1)} \sigma(O_1^{(1)})(1 - \sigma(O_1^{(1)})) O_1^{(0)} \right]$$
$$a_{12}^{(0)} := a_{12}^{(1)} - \alpha \left[2(y - O_0^2) a_{20}^{(1)} \sigma(O_2^{(1)})(1 - \sigma(O_2^{(1)})) O_1^{(0)} \right]$$
$$a_{13}^{(0)} := a_{13}^{(1)} - \alpha \left[2(y - O_0^2) a_{30}^{(1)} \sigma(O_3^{(1)})(1 - \sigma(O_3^{(1)})) O_1^{(0)} \right]$$

$$a_{21}^{(0)} := a_{21}^{(1)} - \alpha \left[2(y - O_0^2) a_{10}^{(1)} \sigma(O_1^{(1)})(1 - \sigma(O_1^{(1)})) O_2^{(0)} \right]$$
$$a_{22}^{(0)} := a_{22}^{(1)} - \alpha \left[2(y - O_0^2) a_{20}^{(1)} \sigma(O_2^{(1)})(1 - \sigma(O_2^{(1)})) O_2^{(0)} \right]$$
$$a_{23}^{(0)} := a_{23}^{(1)} - \alpha \left[2(y - O_0^2) a_{30}^{(1)} \sigma(O_3^{(1)})(1 - \sigma(O_3^{(1)})) O_2^{(0)} \right]$$

Derivatives for the Classification Problem

$$O_k^{(1)} = \sigma \left(\sum_{i=0} O_i^{(0)} \cdot a_{ik}^{(0)} \right) \tag{5}$$

$$X_n^{(2)} = \sum_{k=0} O_k^{(1)} \cdot a_{kn}^{(1)} \tag{6}$$

$$O_n^{(2)} = softmax(X_n^{(2)}, X^{(2)}) \tag{7}$$

General rule for partial derivatives of weights in the second layer:

$$\frac{\partial E(x)}{\partial a_{kn}^{(1)}} = \frac{\partial CE(l, O^{(2)})}{\partial O_n^{(2)}} \frac{\partial O_n^{(2)}}{\partial X_n^{(2}} \frac{\partial X_n^{(2)}}{\partial a_{kn}^{(1)}} \tag{8}$$

$$\frac{\partial E(x)}{\partial a_{kn}^{(1)}} = -l_n \frac{1}{O_n^{(2)}} \frac{\partial O_n^{(2)}}{\partial X_n^{(2)}} \frac{\partial X_n^{(2)}}{\partial a_{kn}^{(1)}}$$

Partial derivative of $O_n^{(2)}$ with respect to $X_n^{(2)}$ can be calculated as below:

$$\frac{\partial log(O_n^{(2)})}{\partial X_n^{(2)}} = \frac{1}{O_n^{(2)}}\frac{\partial log(O_n^{(2)})}{\partial X_n^{(2)}}$$

$$\frac{\partial O_n^{(2)}}{\partial X_n^{(2)}} = O_n^{(2)}\frac{\partial log(O_n^{(2)})}{\partial X_n^{(2)}}$$

$$log(O_n^{(2)}) = log\left(\frac{e^{X_n^{(2)}}}{\sum_{i=0}}e^{X_i^{(2)}}\right) = X_n^{(2)} - log\left(\sum_{i=0} e^{X_i^{(2)}}\right)$$

$$\frac{\partial log(O_n^{(2)})}{\partial X_n^{(2)}} = 1 - \frac{1}{\sum_{i=0} e^{X_i^{(2)}}}e^{X_n^{(2)}}$$

$$\frac{\partial O_n^{(2)}}{\partial X_n^{(2)}} = O_n^{(2)}(1 - O_n^{(2)})$$

Using this equality, the general rule for partial derivative for second layer weights can be calculated.

$$\frac{\partial E(x)}{\partial a_{kn}^{(1)}} = -l_n(1 - O_n^{(2)})O_k^{(1)} \tag{9}$$

All update rules for weights in the second layer can be found using equation 9.

$$a_{00}^{(1)} := a_{00}^{(1)} - l_n(1 - O_0^{(2)})O_0^{(1)}$$
$$a_{10}^{(1)} := a_{10}^{(1)} - l_n(1 - O_0^{(2)})O_1^{(1)}$$
$$a_{20}^{(1)} := a_{20}^{(1)} - l_n(1 - O_0^{(2)})O_2^{(1)}$$
$$a_{30}^{(1)} := a_{30}^{(1)} - l_n(1 - O_0^{(2)})O_3^{(1)}$$

$$a_{01}^{(1)} := a_{01}^{(1)} - l_n(1 - O_1^{(2)})O_0^{(1)}$$
$$a_{11}^{(1)} := a_{11}^{(1)} - l_n(1 - O_1^{(2)})O_1^{(1)}$$
$$a_{21}^{(1)} := a_{21}^{(1)} - l_n(1 - O_1^{(2)})O_2^{(1)}$$
$$a_{31}^{(1)} := a_{31}^{(1)} - l_n(1 - O_1^{(2)})O_3^{(1)}$$

$$a_{02}^{(1)} := a_{02}^{(1)} - l_n(1 - O_2^{(2)})O_0^{(1)}$$
$$a_{12}^{(1)} := a_{12}^{(1)} - l_n(1 - O_2^{(2)})O_1^{(1)}$$
$$a_{20}^{(1)} := a_{22}^{(1)} - l_n(1 - O_2^{(2)})O_2^{(1)}$$
$$a_{30}^{(1)} := a_{32}^{(1)} - l_n(1 - O_2^{(2)})O_3^{(1)}$$