CENG499 Assignment 1

İsmail Karabaş 2375186

October 29, 2023

We know that
Regression architecture calculations

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

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

Classification architecture calculations

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

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

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

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

$\mathrm{softmax}(x, X^{(2)}) = \frac{e^x}{\sum_{s=0}^{S} e^{X_s^{(2)}}}$

Question 1: Regression Update Rules

$a_{ik}^{(0)} = a_{ik}^{(0)} - \alpha \frac{\partial SE(y,O_0^{(2)})}{\partial a_{ik}^{(0)}}$

$a_{k0}^{(1)} = a_{k0}^{(1)} - \alpha \frac{\partial SE(y,O_0^{(2)})}{\partial a_{k0}^{(1)}}$

Derivation:
Given the squared error function for regression:

$SE(y, O_0^{(2)}) = (y - O_0^{(2)})^2$

1. Partial derivative with respect to $a_{k0}^{(1)}$:

Now, applying the chain rule:

$\frac{\partial SE(y,O_0^{(2)})}{\partial a_{k0}^{(1)}} = \frac{\partial SE(y,O_0^{(2)})}{\partial O_0^{(2)}} \cdot \frac{\partial O_0^{(2)}}{\partial a_{k0}^{(1)}}$
$\frac{\partial SE(y,O_0^{(2)})}{\partial a_{k0}^{(1)}} = -2(y - O_0^{(2)}) \cdot \frac{\partial O_0^{(2)}}{\partial a_{k0}^{(1)}}$
lets calculate $\frac{\partial O_0^{(2)}}{\partial a_{k0}^{(1)}}$

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

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

Substitute this back:

$\frac{\partial SE(y,O_0^{(2)})}{\partial a_{k0}^{(1)}} = -2(y - O_0^{(2)})O_k^{(1)}$
So the update rule is :

$$a_{k0}^{(1)} = a_{k0}^{(1)} + 2\alpha(y - O_0^{(2)})O_k^{(1)}$$

2. Partial derivative with respect to $a_{ik}^{(0)}$:

$a_{ik}^{(0)} = a_{ik}^{(0)} - \alpha \frac{\partial SE(y,O_0^{(2)})}{\partial a_{ik}^{(0)}}$

Derivation:
Given the squared error function for regression:

$SE(y, O_0^{(2)}) = (y - O_0^{(2)})^2$

Now, applying the chain rule:

$\frac{\partial SE(y,O_0^{(2)})}{\partial a_{ik}^{(0)}} = \frac{\partial SE(y,O_0^{(2)})}{\partial O_0^{(2)}} \cdot \frac{\partial O_0^{(2)}}{\partial a_{ik}^{(0)}}$
$\frac{\partial SE(y,O_0^{(2)})}{\partial a_{ik}^{(0)}} = -2(y - O_0^{(2)}) \cdot \frac{\partial O_0^{(2)}}{\partial a_{ik}^{(0)}}$
lets calculate $\frac{\partial O_0^{(2)}}{\partial a_{ik}^{(0)}}$

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

$O_0^{(2)} = \sum_{k=0}^{K} \sigma\left(\sum_{i=0}^{n} O_i^{(0)} \cdot a_{ik}^{(0)}\right) \cdot a_{k0}^{(1)}$

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

lets $N = \left(\sum_{i=0}^{n} O_i^{(0)} \cdot a_{ik}^{(0)}\right)$

$\frac{\partial O_0^{(2)}}{\partial a_{ik}^{(0)}} = a_{k0}^{(1)} \frac{\partial O_k^{(1)}}{\partial a_{ik}^{(0)}} = a_{k0}^{(1)} \frac{\partial O_k^{(1)}}{\partial N} \frac{\partial N}{\partial a_{ik}^{(0)}}$

taking the derivative of sigmoid function

$a_{k0}^{(1)} \frac{\partial O_k^{(1)}}{\partial N} \frac{\partial N}{\partial a_{ik}^{(0)}} = a_{k0}^{(1)} N(1-N) \frac{\partial N}{\partial a_{ik}^{(0)}}$

Since the derivative of rest is zero except when it is i and k

$a_{k0}^{(1)} N(1-N) \frac{\partial N}{\partial a_{ik}^{(0)}} = a_{k0}^{(1)} N(1-N) O_i^{(0)}$

Substitute this back:

$\frac{\partial SE(y, O_0^{(2)})}{\partial a_{ik}^{(1)}} = -2(y - O_0^{(2)}) a_{k0}^{(1)} N(1-N) O_i^{(0)}$
So the update rule is :

$$a_{ik}^{(0)} = a_{ik}^{(0)} + 2\alpha(y - O_0^{(2)}) a_{k0}^{(1)} N(1-N) O_i^{(0)}$$

Question 2: Classification Update Rule

We know that

$a_{kn}^{(1)} = a_{kn}^{(1)} - \alpha \frac{\partial CE([l_0, l_1, l_2], \mathbf{O}^{(2)})}{\partial a_{kn}^{(1)}}$

$CE(\mathbf{l}, \mathbf{O}^{(2)}) = -\sum_i l_i \log(O_i^{(2)})$

Calculating the partial derivative using chain rule

$\frac{\partial CE(\mathbf{l}, \mathbf{O}^{(2)})}{\partial a_{kn}^{(1)}} = -\sum_i \frac{l_i}{O_i^{(2)}} \cdot \frac{\partial O_i^{(2)}}{\partial a_{kn}^{(1)}}$

Given:

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

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

$\frac{\partial CE(\mathbf{l}, \mathbf{O}^{(2)})}{\partial a_{kn}^{(1)}} = -\sum_i \frac{l_i}{O_i^{(2)}} \cdot \frac{\partial O_i^{(2)}}{\partial a_{kn}^{(1)}}$

Now, let's find the derivative $\frac{\partial O_i^{(2)}}{\partial a_{kn}^{(1)}}$:

We know that :

$S(\mathbf{x})_i = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$

$\frac{\partial S(\mathbf{x})_i}{\partial x_i} = S(\mathbf{x})_i \cdot (1 - S(\mathbf{x})_i)$

$\frac{\partial S(\mathbf{x})_i}{\partial x_j} = -S(\mathbf{x})_i \cdot S(\mathbf{x})_j$

Lets combine all of these findings

$$\frac{\partial O_i^{(2)}}{\partial a_{kn}^{(1)}} = \frac{\partial O_i^{(2)}}{\partial X_n^{(2)}} \frac{\partial X_n^{(2)}}{\partial a_{kn}^{(1)}} \begin{cases} O_k^{(1)} \cdot (1 - \mathrm{softmax}(X_i^{(2)}, X^{(2)})), & \text{if } i = n \\ -O_k^{(1)} \cdot \mathrm{softmax}(X_i^{(2)}, X^{(2)}) \cdot \mathrm{softmax}(X_n^{(2)}, X^{(2)}), & \text{if } i \neq n \end{cases}$$

Now, substitute this into the partial derivative expression:

$$\frac{\partial CE(\mathbf{l}, \mathbf{O}^{(2)})}{\partial a_{kn}^{(1)}} = -\sum_i \frac{l_i}{O_i^{(2)}} \cdot \frac{\partial O_i^{(2)}}{\partial a_{kn}^{(1)}} = -\sum_i \frac{l_i}{O_i^{(2)}} \cdot \begin{cases} O_k^{(1)} \cdot (1 - \mathrm{softmax}(X_i^{(2)}, X^{(2)})), & \text{if } i = n \\ -O_k^{(1)} \cdot \mathrm{softmax}(X_i^{(2)}, X^{(2)}) \cdot \mathrm{softmax}(X_n^{(2)}, X^{(2)}), & \text{if } i \neq n \end{cases}$$

So the update rule is :

$$a_{kn}^{(1)} = a_{kn}^{(1)} + \alpha \sum_i \frac{l_i}{O_i^{(2)}} \cdot \begin{cases} O_k^{(1)} \cdot (1 - \mathrm{softmax}(X_i^{(2)}, X^{(2)})), & \text{if } i = n \\ -O_k^{(1)} \cdot \mathrm{softmax}(X_i^{(2)}, X^{(2)}) \cdot \mathrm{softmax}(X_n^{(2)}, X^{(2)}), & \text{if } i \neq n \end{cases}$$