SoftArgmax

Heedong Park

July 2022

1 Definition

SoftArgmax($\mathbf{X}$) : $\mathbf{R^n} \longrightarrow \mathbf{R}$
$(x_1, x_2, ..., x_n) \longmapsto SA(x_i)$: Approximate value of i, if $max(X) = x_i$

$$SA(X) = \sum_{i \in n} \frac{e^{x_i}}{\sum\limits_{k \in n} e^{x_k}} i \tag{1}$$

2 Features

• Outputs of SoftArgmax is a index of maxium element of X
(i.e SoftArgmax(X) = i, if max(X) = x_i)

• Differentiable

3 Derivative

$$SA'(X) = \frac{\partial SA}{\partial x_j} = \frac{\partial \sum_{i \in n} \frac{e^{x_i}}{\sum\limits_{k \in n} e^{x_k}} i}{\partial x_j} \tag{2}$$

Then, we can say

$$\frac{\partial \sum_{i \in n} \frac{e^{x_i}}{\sum\limits_{k \in n} e^{x_k}} i}{\partial x_j} = \frac{\sum_{i \in n} \partial \frac{e^{x_i}}{\sum\limits_{k \in n} e^{x_k}} i}{\partial x_j} \tag{3}$$

First, we will focus on

$$SA_i = \frac{e^{x_i}}{\sum\limits_{k \in n} e^{x_k}} i \tag{4}$$

To represent (1) more simple, let

$$g_i = i * e^{x_i}, \quad g_i\prime = (i+1) * e^{x_i} \tag{5}$$

$$h_i = \sum_{k \in n} e^{x_k} = \Sigma, \quad h_i\prime = e^{x_i} \tag{6}$$

Then we can express SA_i as $\frac{g_i}{h_i}$ so that calculating SA'_i, by Quotient Rule

Before differentiating, we have to divide into two cases 1)$i = j$, 2)$i \neq j$

1)i = j

$$SA'_i = \frac{\partial \frac{e^{x_i}}{\sum\limits_{k \in n} e^{x_k}} i}{\partial x_j} = \frac{g'_i * h_i - g_i * h'_i}{h_i^2} \tag{7}$$

$$= \frac{(i+1) * e^{x_i} * \Sigma - i * e^{x_i} * e^{x_j}}{\Sigma^2} = \frac{(i+1) * e^{x_i}}{\Sigma} - \frac{i * e^{x_i} * e^{x_j}}{\Sigma^2} \tag{8}$$

$$= SA_i + \frac{SA_i}{i} - \frac{SA_i^2}{i} = SA_i(1 + \frac{1}{i} - \frac{SA_i}{i}) \tag{9}$$

2)i ≠ j

$$SA'_i = \frac{\partial \frac{e^{x_i}}{\sum\limits_{k \in n} e^{x_k}} i}{\partial x_j} = \frac{g'_i * h_i - g_i * h'_i}{h_i^2} \tag{10}$$

$$= \frac{-i * e^{x_i} * e^{x_j}}{\Sigma^2} = -\frac{(i * e^{x_i}) * (i * e^{x_j})}{i * \Sigma * \Sigma} \tag{11}$$

$$= -(\frac{1}{i} * SA_i * SA_j) \tag{12}$$

Now everything is ready
Remind formula (3), then we can say

$$SA'(X) = -\sum_{i \neq j} \frac{SA_i * SA_j}{i} + SA_j(1 + \frac{1}{j} - \frac{SA_j}{j}) \tag{13}$$

$$\therefore -\sum_{i}^{n} \frac{SA_i * SA_j}{j} + SA_j(1 + \frac{1}{j}) \tag{14}$$

(Note that, j is index of $x_j \in X$)

4 Interpretation

Let's see (14) again. we can modify this formula slightly

$$-\frac{SA_j}{j}\sum_i^n SA_i + SA_j + \frac{SA_j}{j} \tag{15}$$

Then $\frac{SA_j}{j}$ is equal to Softmax-function's value and it is obtained during "Foward". It seems to be advantageous in calculationg and still looks as simple as sigmoid, softmax derivative forms.
If we write S_j as a softmax(x_j),

$$-S_j\sum_i^n SA_i + SA_j + S_j \tag{16}$$