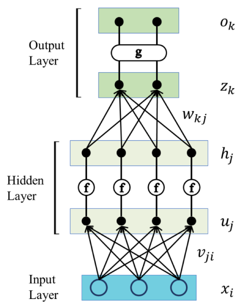
2018170622 김지은 HW #2

Problem 1. (60points)

Consider a neural network which receives an input $\boldsymbol{x} = [\,x_1 \quad x_2 \quad x_3\,]^t$ and generates an output $\boldsymbol{o} = [\,o_1 \quad o_2\,]^t$. It consists of one input layer, one hidden layer, and one output layer. The hidden layer contains 4 neurons and the output layer has 2 neurons. The input $\boldsymbol{x}$ is transformed by a set of weights $\{v_{ji} \mid j = 1,2,3,4,\ i = 1,2,3\}$ and the output of the hidden layer is transformed by weights $\{w_{kj} \mid k = 1,2,\ j = 1,2,3,4\}$. f is a *ELU* activation function:

$$\mathrm{ELU}(u) = \begin{cases} u & \geq 0 \\ e^u - 1 & < 0 \end{cases}.$$

g is a *Softmax* activation function.



1) **(10points)** Derive the expressions for u_j, h_j, z_k, and o_k using the input, activation function, and weights.

$$u_j = \sum_i v_{ji} x_i = \sum_{i=1}^{3} v_{ji} x_i = v_{j1}x_1 + v_{j2}x_2 + v_{j3}x_3$$

$$h_j = f(u_j) = \mathrm{ELU}(u_j) = \mathrm{ELU}(v_{j1}x_1 + v_{j2}x_2 + v_{j3}x_3)$$

$$z_k = \sum_j w_{kj} h_j = \sum_j w_{kj}\,\mathrm{ELU}(u_j) = \sum_{j=1}^{4} w_{kj}\,\mathrm{ELU}(v_{j1}x_1 + v_{j2}x_2 + v_{j3}x_3)$$

$$o_k = g(z_k) = \frac{e^{z_k}}{e^{z_1} + e^{z_2}} = \frac{e^{\sum_{j=1}^{4} w_{kj}\mathrm{ELU}(\sum_{i=1}^{3} v_{ji}x_i)}}{e^{\sum_{j=1}^{4} w_{1j}\mathrm{ELU}(\sum_{i=1}^{3} v_{ji}x_i)} + e^{\sum_{j=1}^{4} w_{2j}\mathrm{ELU}(\sum_{i=1}^{3} v_{ji}x_i)}}$$

2) **(10points)** Suppose that $\boldsymbol{x} = [\,1.0 \quad 1.5 \quad -0.5\,]^t$ and the ground truth $o^* = [\,1 \quad 0\,]^t$. And, the loss function $\mathcal{L}$ and the initial weights are given as follows:

$$\mathcal{L} = -\sum_{k=1}^{2} o^*_k \log o_k$$

$$v = \begin{bmatrix} -1.0 & 2.0 & 0.0 \\ 0.5 & -1.0 & 1.0 \\ 1.5 & 2.0 & -0.4 \\ 0.2 & 0.1 & 0.3 \end{bmatrix}$$

$$w = \begin{bmatrix} 1.0 & 1.0 & 0.5 & 0.5 \\ 0.3 & 0.4 & 0.5 & 0.1 \end{bmatrix}.$$

Compute u_j, h_j, z_k, o_k, and $\mathcal{L}$ ($j = 1,2,3,4$ and $k = 1,2$).

$$u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \vec{v} \cdot \vec{x} = \begin{bmatrix} -1.0 & 2.0 & 0.0 \\ 0.5 & -1.0 & 1.0 \\ 1.5 & 2.0 & -0.4 \\ 0.2 & 0.1 & 0.3 \end{bmatrix} \begin{bmatrix} 1.0 \\ 1.5 \\ -0.5 \end{bmatrix} = [2, -1.5, 4.7, 0.2]^T$$

$$h = \begin{bmatrix} \mathrm{ELU}(u_1) \\ \mathrm{ELU}(u_2) \\ \mathrm{ELU}(u_3) \\ \mathrm{ELU}(u_4) \end{bmatrix} = [2, -0.7769, 4.7, 0.2]^T$$

$$z = \vec{w} \cdot h = \begin{bmatrix} 1.0 & 1.0 & 0.5 & 0.5 \\ 1.3 & 0.4 & 0.5 & 0.1 \end{bmatrix} \begin{bmatrix} 2 \\ -0.7769 \\ 4.7 \\ 0.2 \end{bmatrix} = [3.693, 2.659]^T$$

$$O = g(z) = \begin{bmatrix} g(z_1) \\ g(z_2) \end{bmatrix} = \begin{bmatrix} 0.7338 \\ 0.2662 \end{bmatrix}$$

$$\mathcal{L} = -(o_1^* \log o_1 + o_2^* \log o_2) = -(1 \times \log(0.7338)) = 0.1344$$

3) **(10points)** Derive the expressions for $\frac{\partial \mathcal{L}}{\partial w_{kj}}$ and $\frac{\partial \mathcal{L}}{\partial v_{ji}}$. (Hint: $\frac{\partial \mathcal{L}}{\partial z_k} = o_k - o^*{}_k$)

$$\frac{\partial \mathcal{L}}{\partial w_{kj}} = \underbrace{\frac{\partial \mathcal{L}}{\partial o_k}\frac{\partial o_k}{\partial z_k}}_{o_k - o_k^*}\underbrace{\frac{\partial z_k}{\partial w_{kj}}}_{h_j} = (o_k - o_k^*)h_j$$

$$\frac{\partial \mathcal{L}}{\partial v_{ji}} = \frac{\partial \mathcal{L}}{\partial h_j} \cdot \frac{\partial h_j}{\partial v_{ji}} = \frac{\partial \mathcal{L}}{\partial h_j} \cdot \underbrace{\frac{\partial h_j}{\partial u_j}}_{f'(u_j)} \cdot \underbrace{\frac{\partial u_j}{\partial v_{ji}}}_{x_i} = f'(u_j)\,x_i \sum_k (o_k - o_k^*) w_{kj}$$

$$\frac{\partial \mathcal{L}}{\partial h_j} = \sum_k \frac{\partial \mathcal{L}}{\partial o_k} \cdot \frac{\partial o_k}{\partial z_k} \cdot \underbrace{\frac{\partial z_k}{\partial h_j}}_{w_{kj}} = \sum_k (o_k - o_k^*) \cdot w_{kj}$$

$(o_1 - o_1^*) w_{1j}$
$+ (o_2 - o_2^*) w_{2j}$

$$f'(u) \begin{cases} 1 & (u \geq 0) \\ e^u & (u < 0) \end{cases}$$

$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}$ $\begin{bmatrix} 0 \\ \vdots \\ 3 \end{bmatrix}$ $f'(u_1) \cdot x_1$

4) **(15points)** Given $\boldsymbol{x} = [\,1.0 \quad 1.5 \quad -0.5\,]^t$, compute the values of $\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}}$ and $\frac{\partial \mathcal{L}}{\partial \boldsymbol{v}}$.

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial w_{11}} & \frac{\partial \mathcal{L}}{\partial w_{12}} & \frac{\partial \mathcal{L}}{\partial w_{13}} & \frac{\partial \mathcal{L}}{\partial w_{14}} \\ \frac{\partial \mathcal{L}}{\partial w_{21}} & \frac{\partial \mathcal{L}}{\partial w_{22}} & \frac{\partial \mathcal{L}}{\partial w_{23}} & \frac{\partial \mathcal{L}}{\partial w_{24}} \end{bmatrix}, \quad \frac{\partial \mathcal{L}}{\partial \boldsymbol{v}} = \underset{4\times 3}{\begin{bmatrix} \frac{\partial \mathcal{L}}{\partial v_{11}} & \frac{\partial \mathcal{L}}{\partial v_{12}} & \frac{\partial \mathcal{L}}{\partial v_{13}} \\ \frac{\partial \mathcal{L}}{\partial v_{21}} & \frac{\partial \mathcal{L}}{\partial v_{22}} & \frac{\partial \mathcal{L}}{\partial v_{23}} \\ \frac{\partial \mathcal{L}}{\partial v_{31}} & \frac{\partial \mathcal{L}}{\partial v_{32}} & \frac{\partial \mathcal{L}}{\partial v_{33}} \\ \frac{\partial \mathcal{L}}{\partial v_{41}} & \frac{\partial \mathcal{L}}{\partial v_{42}} & \frac{\partial \mathcal{L}}{\partial v_{43}} \end{bmatrix}}$$

given $\vec{x}$에서 $\vec{O}$와 h는

$$\vec{O} = [0.7338 \quad 0.2662]^T$$

$$h = [2, -0.7769, 4.7, 0.2]^T$$

$$\therefore \vec{O} - \vec{O}^* = [-0.2662 \quad 0.2662]^T$$

$$\therefore \frac{\partial \mathcal{L}}{\partial \vec{w}} = [-0.9713, 0.0495] \cdot [2, -0.7769, 4.7, 0.2]^T$$

$$= \begin{bmatrix} -0.5324 & 0.2068 & -1.251 & -0.0532 \\ 0.5324 & -0.2068 & 1.251 & 0.0532 \end{bmatrix}$$

$$\therefore \frac{\partial \mathcal{L}}{\partial \vec{v}} = \begin{bmatrix} -0.1863 & -0.2795 & 0.0932 \\ -0.0356 & -0.0535 & 0.0178 \\ 0 & 0 & 0 \\ -0.1065 & -0.1597 & 0.0532 \end{bmatrix}$$

5) **(15points)** Using the values of $\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}}$ and $\frac{\partial \mathcal{L}}{\partial \boldsymbol{v}}$ from 4) and a learning rate $\eta = 0.1$, update the weights $\boldsymbol{v}$ and $\boldsymbol{W}$ via a gradient descent update rule.

$$\vec{w}_{new} = \vec{w}_{old} - 0.1 \nabla \mathcal{L}(\vec{w})$$

$$w_i^{new} = w_i^{old} - 0.1 \frac{\partial \mathcal{L}}{\partial w_i}$$

$$= \begin{bmatrix} 1.0532 & 0.9793 & 0.651 & 0.5053 \\ 0.2468 & 0.4207 & 0.3747 & 0.0947 \end{bmatrix}$$