EECE7398 – Homework 2

1.

SoftMax probability for the correct class when $j = y$: $\quad P_y = \left(\frac{e^{z_y}}{\sum_{k=1}^{C} e^{z_k}} \right)$

$$
\begin{aligned}
\frac{\partial L}{\partial z_y} &= \frac{\partial}{\partial z_y} \left(-\log \left(\frac{e^{z_y}}{\sum_{k=1}^{C} e^{z_k}} \right) \right) \\
&= \frac{\partial}{\partial z_y} \left(-\log(e^{z_y}) + \log \left(\sum_{k=1}^{C} e^{z_k} \right) \right) \\
&= \frac{\partial}{\partial z_y} \left(-z_y + \log \left(\sum_{k=1}^{C} e^{z_k} \right) \right) \\
&= -1 + \frac{\partial}{\partial z_y} \left(\log \left(\sum_{k=1}^{C} e^{z_k} \right) \right) \\
&= -1 + \frac{1}{\sum_{k=1}^{C} e^{z_k}} \cdot \frac{\partial}{\partial z_y} \left(\sum_{k=1}^{C} e^{z_k} \right) \\
&= -1 + \frac{1}{\sum_{k=1}^{C} e^{z_k}} \cdot e^{z_y} \\
&= -1 + \frac{e^{z_y}}{\sum_{k=1}^{C} e^{z_k}} \\
&= -1 + P_y
\end{aligned}
$$

SoftMax probability for incorrect class when $j \neq y$: $\quad P_j = \left(\frac{e^{z_j}}{\sum_{k=1}^{C} e^{z_k}} \right)$

$$
\begin{aligned}
\frac{\partial L}{\partial z_j} &= \frac{\partial}{\partial z_j}\left(-\log\left(\frac{e^{z_y}}{\sum_{k=1}^{C} e^{z_k}}\right)\right) \\
&= \frac{\partial}{\partial z_j}\left(-\log(e^{z_y}) + \log\left(\sum_{k=1}^{C} e^{z_k}\right)\right) \\
&= \frac{\partial}{\partial z_j}\left(-z_y + \log\left(\sum_{k=1}^{C} e^{z_k}\right)\right) \\
&= 0 + \frac{\partial}{\partial z_j}\left(\log\left(\sum_{k=1}^{C} e^{z_k}\right)\right) \\
&= \frac{1}{\sum_{k=1}^{C} e^{z_k}} \cdot \frac{\partial}{\partial z_j}\left(\sum_{k=1}^{C} e^{z_k}\right) \\
&= \frac{1}{\sum_{k=1}^{C} e^{z_k}} \cdot e^{z_j} \\
&= \frac{e^{z_j}}{\sum_{k=1}^{C} e^{z_k}} \\
&= P_j
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial z_j}{\partial W_j} &= \frac{\partial}{\partial W_j}(W_j \cdot x) \\
&= x
\end{aligned}
$$

Gradient for the correct class $j = y:$ $\quad \nabla_W L = \frac{\partial L}{\partial W_j} = \frac{\partial L}{\partial z_j} \cdot \frac{\partial z_j}{\partial W_j} = (P_y - 1) \cdot x$

Gradient for the incorrect class $j \neq y:$ $\quad \nabla_W L = \frac{\partial L}{\partial W_j} = \frac{\partial L}{\partial z_j} \cdot \frac{\partial z_j}{\partial W_j} = P_j \cdot x$

Gradient for the correct class $j = y$ in matrix form: $\quad \nabla_W L = (P_y - Y_y) \cdot x^T$

Gradient for the incorrect class $j \neq y$ in matrix form: $\quad \nabla_W L = (P_j - Y_j) \cdot x^T = (P_j - 0) \cdot x^T = P_j \cdot x^T$

Gradient general matrix form: $\quad \nabla_W L = (P - Y) \cdot x^T$

$$
x_1 = \begin{bmatrix} 1.52 \\ 2.63 \\ 5.37 \\ 4.94 \end{bmatrix}, \quad y = \text{cat}
$$

$$
z_1 = W \cdot x_1 = \begin{bmatrix} -0.57 & 1.24 & -3.37 & 6.43 \\ -5.53 & -1.13 & -8.05 & 3.21 \\ 4.23 & 0.98 & -2.53 & -7.67 \\ -2.31 & -1.84 & 6.93 & -8.66 \end{bmatrix} \cdot \begin{bmatrix} 1.52 \\ 2.63 \\ 5.37 \\ 4.94 \end{bmatrix} = \begin{bmatrix} 16.0621 \\ -38.7486 \\ -42.4689 \\ -13.9217 \end{bmatrix}
$$

$$P_j = \left(\frac{e^{z_j}}{\sum_{k=1}^{C} e^{z_k}} \right)$$

$$P_{\text{cat}} = \left(\frac{e^{z_{\text{cat}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{16.0621}}{e^{16.0621} + e^{-38.7486} + e^{-42.4689} + e^{-13.9217}} \right) = 1$$

$$P_{\text{dog}} = \left(\frac{e^{z_{\text{dog}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{-38.7486}}{e^{16.0621} + e^{-38.7486} + e^{-42.4689} + e^{-13.9217}} \right) \approx 1.38 \times 10^{-24}$$

$$P_{\text{cow}} = \left(\frac{e^{z_{\text{cow}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{-42.4689}}{e^{16.0621} + e^{-38.7486} + e^{-42.4689} + e^{-13.9217}} \right) \approx 4.05 \times 10^{-25}$$

$$P_{\text{horse}} = \left(\frac{e^{z_{\text{horse}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{-13.9217}}{e^{16.0621} + e^{-38.7486} + e^{-42.4689} + e^{-13.9217}} \right) \approx 9.4 \times 10^{-14}$$

$$P = \begin{bmatrix} P_{\text{cat}} \\ P_{\text{dog}} \\ P_{\text{cow}} \\ P_{\text{horse}} \end{bmatrix} = \begin{bmatrix} 1 \\ 1.38 \times 10^{-24} \\ 4.05 \times 10^{-25} \\ 9.4 \times 10^{-14} \end{bmatrix}$$

$$Y = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$P - Y = \begin{bmatrix} 1 \\ 1.38 \times 10^{-24} \\ 4.05 \times 10^{-25} \\ 9.4 \times 10^{-14} \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1.38 \times 10^{-24} \\ 4.05 \times 10^{-25} \\ 9.4 \times 10^{-14} \end{bmatrix}$$

$$x^T = \begin{bmatrix} 1.52 & 2.63 & 5.37 & 4.94 \end{bmatrix}$$

$$\begin{aligned} \nabla_W L_1 &= (P - Y) \cdot x^T \\ &= \begin{bmatrix} 0 \\ 1.38 \times 10^{-24} \\ 4.05 \times 10^{-25} \\ 9.4 \times 10^{-14} \end{bmatrix} \cdot \begin{bmatrix} 1.52 & 2.63 & 5.37 & 4.94 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 2.10 \times 10^{-24} & 3.63 \times 10^{-24} & 7.41 \times 10^{-24} & 6.82 \times 10^{-24} \\ 6.16 \times 10^{-25} & 1.07 \times 10^{-24} & 2.17 \times 10^{-24} & 2.00 \times 10^{-24} \\ 1.43 \times 10^{-13} & 2.47 \times 10^{-13} & 5.05 \times 10^{-13} & 4.64 \times 10^{-13} \end{bmatrix} \end{aligned}$$

$$x_2 = \begin{bmatrix} 8.87 \\ 1.25 \\ 4.49 \\ 0.12 \end{bmatrix}, \quad y = \text{cat}$$

$$z_2 = W \cdot x_2 = \begin{bmatrix} -0.57 & 1.24 & -3.37 & 6.43 \\ -5.53 & -1.13 & -8.05 & 3.21 \\ 4.23 & 0.98 & -2.53 & -7.67 \\ -2.31 & -1.84 & 6.93 & -8.66 \end{bmatrix} \cdot \begin{bmatrix} 8.87 \\ 1.25 \\ 4.49 \\ 0.12 \end{bmatrix} = \begin{bmatrix} -17.8656 \\ -86.2379 \\ 26.466 \\ 7.2968 \end{bmatrix}$$

$$P_j = \left(\frac{e^{z_j}}{\sum_{k=1}^{C} e^{z_k}} \right)$$

$$P_{\text{cat}} = \left(\frac{e^{z_{\text{cat}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{-17.8656}}{e^{-17.8656} + e^{-86.2379} + e^{26.466} + e^{7.2968}} \right) = 7.04 \times 10^{-20}$$

$$P_{\text{dog}} = \left(\frac{e^{z_{\text{dog}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{-86.2379}}{e^{-17.8656} + e^{-86.2379} + e^{26.466} + e^{7.2968}} \right) \approx 0$$

$$P_{\text{cow}} = \left(\frac{e^{z_{\text{cow}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{26.466}}{e^{-17.8656} + e^{-86.2379} + e^{26.466} + e^{7.2968}} \right) = 1$$

$$P_{\text{horse}} = \left(\frac{e^{z_{\text{horse}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{7.2968}}{e^{-17.8656} + e^{-86.2379} + e^{26.466} + e^{7.2968}} \right) = 5.96 \times 10^{-9}$$

$$P = \begin{bmatrix} P_{\text{cat}} \\ P_{\text{dog}} \\ P_{\text{cow}} \\ P_{\text{horse}} \end{bmatrix} = \begin{bmatrix} 7.04 \times 10^{-20} \\ 0 \\ 1 \\ 5.96 \times 10^{-9} \end{bmatrix}$$

$$Y = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$P - Y = \begin{bmatrix} 7.04 \times 10^{-20} \\ 0 \\ 1 \\ 5.96 \times 10^{-9} \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 5.96 \times 10^{-9} \end{bmatrix}$$

$$x^T = \begin{bmatrix} 8.87 & 1.25 & 4.49 & 0.12 \end{bmatrix}$$

$$\begin{aligned} \nabla_W L_2 &= (P - Y) \cdot x^T \\ &= \begin{bmatrix} -1 \\ 0 \\ 1 \\ 5.96 \times 10^{-9} \end{bmatrix} \cdot \begin{bmatrix} 8.87 & 1.25 & 4.49 & 0.12 \end{bmatrix} = \begin{bmatrix} -8.87 & -1.25 & -4.49 & -0.12 \\ 0 & 0 & 0 & 0 \\ 8.87 & 1.25 & 4.49 & 0.12 \\ 5.29 \times 10^{-8} & 7.45 \times 10^{-9} & 2.67 \times 10^{-8} & 7.15 \times 10^{-10} \end{bmatrix} \end{aligned}$$

$$x_3 = \begin{bmatrix} 3.22 \\ 4.63 \\ 3.55 \\ 5.41 \end{bmatrix}, \quad y = \text{dog}$$

$$z_3 = W \cdot x_3 = \begin{bmatrix} -0.57 & 1.24 & -3.37 & 6.43 \\ -5.53 & -1.13 & -8.05 & 3.21 \\ 4.23 & 0.98 & -2.53 & -7.67 \\ -2.31 & -1.84 & 6.93 & -8.66 \end{bmatrix} \cdot \begin{bmatrix} 3.22 \\ 4.63 \\ 3.55 \\ 5.41 \end{bmatrix} = \begin{bmatrix} 26.7286 \\ -34.2499 \\ -32.3182 \\ -38.2065 \end{bmatrix}$$

$$P_j = \left(\frac{e^{z_j}}{\sum_{k=1}^{C} e^{z_k}} \right)$$

$$P_{\text{cat}} = \left(\frac{e^{z_{\text{cat}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{26.7286}}{e^{26.7286} + e^{-34.2499} + e^{-32.3182} + e^{-38.2065}} \right) = 1$$

$$P_{\text{dog}} = \left(\frac{e^{z_{\text{dog}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{-34.2499}}{e^{26.7286} + e^{-34.2499} + e^{-32.3182} + e^{-38.2065}} \right) = 3.08 \times 10^{-27}$$

$$P_{\text{cow}} = \left(\frac{e^{z_{\text{cow}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{-32.3182}}{e^{26.7286} + e^{-34.2499} + e^{-32.3182} + e^{-38.2065}} \right) = 1.81 \times 10^{-26}$$

$$P_{\text{horse}} = \left(\frac{e^{z_{\text{horse}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{-38.2065}}{e^{26.7286} + e^{-34.2499} + e^{-32.3182} + e^{-38.2065}} \right) = 5.38 \times 10^{-29}$$

$$P = \begin{bmatrix} P_{\text{cat}} \\ P_{\text{dog}} \\ P_{\text{cow}} \\ P_{\text{horse}} \end{bmatrix} = \begin{bmatrix} 1 \\ 3.08 \times 10^{-27} \\ 1.81 \times 10^{-26} \\ 5.38 \times 10^{-29} \end{bmatrix}$$

$$Y = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$P - Y = \begin{bmatrix} 1 \\ 3.08 \times 10^{-27} \\ 1.81 \times 10^{-26} \\ 5.38 \times 10^{-29} \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1.81 \times 10^{-26} \\ 5.38 \times 10^{-29} \end{bmatrix}$$

$$x^T = \begin{bmatrix} 3.22 & 4.63 & 3.55 & 5.41 \end{bmatrix}$$

$$\begin{aligned} \nabla_W L_3 &= (P - Y) \cdot x^T \\ &= \begin{bmatrix} 1 \\ -1 \\ 1.81 \times 10^{-26} \\ 5.38 \times 10^{-29} \end{bmatrix} \cdot \begin{bmatrix} 3.22 & 4.63 & 3.55 & 5.41 \end{bmatrix} = \begin{bmatrix} 3.22 & 4.63 & 3.55 & 5.41 \\ -3.22 & -4.63 & -3.55 & -5.41 \\ 5.83 \times 10^{-26} & 8.38 \times 10^{-26} & 6.43 \times 10^{-26} & 9.79 \times 10^{-26} \\ 1.73 \times 10^{-28} & 2.49 \times 10^{-28} & 1.91 \times 10^{-28} & 2.91 \times 10^{-28} \end{bmatrix} \end{aligned}$$

$$x_4 = \begin{bmatrix} 1.38 \\ 0.63 \\ 2.90 \\ 8.52 \end{bmatrix}, \quad y = \text{horse}$$

$$z_4 = W \cdot x_4 = \begin{bmatrix} -0.57 & 1.24 & -3.37 & 6.43 \\ -5.53 & -1.13 & -8.05 & 3.21 \\ 4.23 & 0.98 & -2.53 & -7.67 \\ -2.31 & -1.84 & 6.93 & -8.66 \end{bmatrix} \cdot \begin{bmatrix} 1.38 \\ 0.63 \\ 2.90 \\ 8.52 \end{bmatrix} = \begin{bmatrix} 45.0052 \\ -4.3391 \\ -66.2306 \\ -58.0332 \end{bmatrix}$$

$$P_j = \left(\frac{e^{z_j}}{\sum_{k=1}^{C} e^{z_k}} \right)$$

$$P_{\text{cat}} = \left(\frac{e^{z_{\text{cat}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{45.0052}}{e^{45.0052} + e^{-4.3391} + e^{-66.2306} + e^{-58.0332}} \right) = 1$$

$$P_{\text{dog}} = \left(\frac{e^{z_{\text{dog}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{-4.3391}}{e^{45.0052} + e^{-4.3391} + e^{-66.2306} + e^{-58.0332}} \right) = 3.71 \times 10^{-22}$$

$$P_{\text{cow}} = \left(\frac{e^{z_{\text{cow}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{-66.2306}}{e^{45.0052} + e^{-4.3391} + e^{-66.2306} + e^{-58.0332}} \right) = 4.90 \times 10^{-49}$$

$$P_{\text{horse}} = \left(\frac{e^{z_{\text{horse}}}}{\sum_{k \in \{\text{cat, dog, cow, horse}\}} e^{z_k}} \right) = \left(\frac{e^{-58.0332}}{e^{45.0052} + e^{-4.3391} + e^{-66.2306} + e^{-58.0332}} \right) = 1.78 \times 10^{-45}$$

$$P = \begin{bmatrix} P_{\text{cat}} \\ P_{\text{dog}} \\ P_{\text{cow}} \\ P_{\text{horse}} \end{bmatrix} = \begin{bmatrix} 1 \\ 3.71 \times 10^{-22} \\ 4.90 \times 10^{-49} \\ 1.78 \times 10^{-45} \end{bmatrix}$$

$$Y = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$P - Y = \begin{bmatrix} 1 \\ 3.71 \times 10^{-22} \\ 4.90 \times 10^{-49} \\ 1.78 \times 10^{-45} \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3.71 \times 10^{-22} \\ 4.90 \times 10^{-49} \\ -1 \end{bmatrix}$$

$$x^T = \begin{bmatrix} 1.38 & 0.63 & 2.90 & 8.52 \end{bmatrix}$$

$$\begin{aligned} \nabla_W L_4 &= (P - Y) \cdot x^T \\ &= \begin{bmatrix} 1 \\ 4.15 \times 10^{-21} \\ 1.4 \times 10^{-48} \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 1.38 & 0.63 & 2.90 & 8.52 \end{bmatrix} = \begin{bmatrix} 1.38 & 0.63 & 2.90 & 8.52 \\ 5.12 \times 10^{-22} & 2.34 \times 10^{-22} & 1.07 \times 10^{-21} & 3.16 \times 10^{-21} \\ 6.77 \times 10^{-49} & 3.09 \times 10^{-49} & 1.42 \times 10^{-48} & 4.18 \times 10^{-48} \\ -1.38 & -0.63 & -2.90 & -8.52 \end{bmatrix} \end{aligned}$$

2.

$$\nabla_W L = \nabla_W L_1 + \nabla_W L_2 + \nabla_W L_3 + \nabla_W L_4$$

$$\nabla_W L = \begin{bmatrix} -4.27 & 4.01 & 1.96 & 13.81 \\ -3.22 & -4.63 & -3.55 & -5.41 \\ 8.87 & 1.25 & 4.49 & 0.12 \\ -1.38 & -0.63 & -2.90 & -8.52 \end{bmatrix}$$

$$\begin{aligned} \Delta W &= \eta \cdot \nabla_W L \\ &= 0.2 \cdot \begin{bmatrix} -4.27 & 4.01 & 1.96 & 13.81 \\ -3.22 & -4.63 & -3.55 & -5.41 \\ 8.87 & 1.25 & 4.49 & 0.12 \\ -1.38 & -0.63 & -2.90 & -8.52 \end{bmatrix} \\ &= \begin{bmatrix} -0.854 & 0.802 & 0.392 & 2.762 \\ -0.644 & -0.926 & -0.710 & -1.082 \\ 1.774 & 0.250 & 0.898 & 0.024 \\ -0.276 & -0.126 & -0.580 & -1.704 \end{bmatrix} \end{aligned}$$

$$W_{\text{new}} = W_{\text{old}} - \Delta W$$

$$\begin{aligned} W_{\text{new}} &= \begin{bmatrix} -0.57 & 1.24 & -3.37 & 6.43 \\ -5.53 & -1.13 & -8.05 & 3.21 \\ 4.23 & 0.98 & -2.53 & -7.67 \\ -2.31 & -1.84 & 6.93 & -8.66 \end{bmatrix} - \begin{bmatrix} -0.854 & 0.802 & 0.392 & 2.762 \\ -0.644 & -0.926 & -0.710 & -1.082 \\ 1.774 & 0.250 & 0.898 & 0.024 \\ -0.276 & -0.126 & -0.580 & -1.704 \end{bmatrix} \\ &= \begin{bmatrix} 0.284 & 0.438 & -3.762 & 3.668 \\ -4.886 & -0.204 & -7.340 & 4.292 \\ 2.456 & 0.730 & -3.428 & -7.694 \\ -2.034 & -1.714 & 7.510 & -6.956 \end{bmatrix} \end{aligned}$$

3.

$$\begin{aligned} w_1 &= -1.7 \\ w_2 &= 0.1 \\ w_3 &= -0.6 \\ w_4 &= -1.8 \\ w_5 &= -0.2 \\ w_6 &= 0.5 \\ x_1 &= -0.3 \\ x_2 &= 4.9 \\ x_3 &= 1.1 \\ x_4 &= -2.7 \end{aligned}$$

$$\begin{aligned} s_1 &= w_1 \cdot x_1 + w_2 \cdot x_2 \\ &= -1.7 \cdot (-0.3) + 0.1 \cdot 4.9 = 1 \end{aligned}$$

$$
\begin{aligned}
s_2 &= w_3 \cdot x_3 + w_4 \cdot x_4 \\
&= -0.6 \cdot 1.1 + -1.8 \cdot -2.7 = 4.2
\end{aligned}
$$

$$
\begin{aligned}
h_1 = \sigma(s_1) &= \frac{1}{1+e^{-s_1}} \\
&= \frac{1}{1+e^{-1}} \approx 0.731
\end{aligned}
$$

$$
\begin{aligned}
h_2 = \sigma(s_2) &= \frac{1}{1+e^{-s_2}} \\
&= \frac{1}{1+e^{-4.2}} \approx 0.985
\end{aligned}
$$

$$
\begin{aligned}
s_3 &= w_5 \cdot h_1 + w_6 \cdot h_2 \\
&= -0.2 \cdot 0.731 + 0.5 \cdot 0.985 \approx 0.3463
\end{aligned}
$$

$$
\begin{aligned}
\hat{y} = \sigma(s_3) &= \frac{1}{1+e^{-s_3}} \\
&= \frac{1}{1+e^{-0.3463}} \approx 0.586
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial w_1} &= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial s_3} \cdot \frac{\partial s_3}{\partial h_1} \cdot \frac{\partial h_1}{\partial s_1} \cdot \frac{\partial s_1}{\partial w_1} \\
&= \frac{\partial}{\partial \hat{y}} \left(\|y - \hat{y}\|^2 \right) \cdot \frac{\partial}{\partial s_3} \left(\sigma(s_3) \right) \cdot \frac{\partial}{\partial h_1} \left(w_5 \cdot h_1 + w_6 \cdot h_2 \right) \cdot \frac{\partial}{\partial s_1} \left(\sigma(s_1) \right) \cdot \frac{\partial}{\partial w_1} \left(w_1 \cdot x_1 + w_2 \cdot x_2 \right) \\
&= 2(\hat{y} - y) \cdot \sigma(s_3)(1 - \sigma(s_3)) \cdot w_5 \cdot \sigma(s_1)(1 - \sigma(s_1)) \cdot x_1 \quad (As \frac{d}{dx}\sigma(x) = \sigma(x)(1 - \sigma(x))) \\
&= 2(\hat{y} - y) \cdot \hat{y}(1 - \hat{y}) \cdot w_5 \cdot h_1(1 - h_1) \cdot x_1 \\
&= 2(0.586 - 0.7) \cdot 0.586(1 - 0.586) \cdot -0.2 \cdot 0.731(1 - 0.731) \cdot -0.3 \\
&\approx -0.00065
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial w_2} &= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial s_3} \cdot \frac{\partial s_3}{\partial h_1} \cdot \frac{\partial h_1}{\partial s_1} \cdot \frac{\partial s_1}{\partial w_2} \\
&= \frac{\partial}{\partial \hat{y}} \left(\|y - \hat{y}\|^2 \right) \cdot \frac{\partial}{\partial s_3} \left(\sigma(s_3) \right) \cdot \frac{\partial}{\partial h_1} \left(w_5 \cdot h_1 + w_6 \cdot h_2 \right) \cdot \frac{\partial}{\partial s_1} \left(\sigma(s_1) \right) \cdot \frac{\partial}{\partial w_2} \left(w_1 \cdot x_1 + w_2 \cdot x_2 \right) \\
&= 2(\hat{y} - y) \cdot \sigma(s_3)(1 - \sigma(s_3)) \cdot w_5 \cdot \sigma(s_1)(1 - \sigma(s_1)) \cdot x_2 \\
&= 2(\hat{y} - y) \cdot \hat{y}(1 - \hat{y}) \cdot w_5 \cdot h_1(1 - h_1) \cdot x_2 \\
&= 2(0.586 - 0.7) \cdot 0.586(1 - 0.586) \cdot -0.2 \cdot 0.731(1 - 0.731) \cdot 4.9 \\
&\approx 0.01066
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial w_3} &= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial s_3} \cdot \frac{\partial s_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial s_2} \cdot \frac{\partial s_2}{\partial w_3} \\
&= \frac{\partial}{\partial \hat{y}} \left(\|y - \hat{y}\|^2 \right) \cdot \frac{\partial}{\partial s_3} \left(\sigma(s_3) \right) \cdot \frac{\partial}{\partial h_2} \left(w_5 \cdot h_1 + w_6 \cdot h_2 \right) \cdot \frac{\partial}{\partial s_2} \left(\sigma(s_2) \right) \cdot \frac{\partial}{\partial w_3} \left(w_3 \cdot x_3 + w_4 \cdot x_4 \right) \\
&= 2(\hat{y} - y) \cdot \sigma(s_3)(1 - \sigma(s_3)) \cdot w_6 \cdot \sigma(s_2)(1 - \sigma(s_2)) \cdot x_3 \\
&= 2(\hat{y} - y) \cdot \hat{y}(1 - \hat{y}) \cdot w_6 \cdot h_2(1 - h_2) \cdot x_3 \\
&= 2(0.586 - 0.7) \cdot 0.586(1 - 0.586) \cdot 0.5 \cdot 0.985(1 - 0.985) \cdot 1.1 \\
&\approx -0.00045
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial w_4} &= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial s_3} \cdot \frac{\partial s_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial s_2} \cdot \frac{\partial s_2}{\partial w_4} \\
&= \frac{\partial}{\partial \hat{y}} \left(\|y - \hat{y}\|^2 \right) \cdot \frac{\partial}{\partial s_3} \left(\sigma(s_3) \right) \cdot \frac{\partial}{\partial h_2} \left(w_5 \cdot h_1 + w_6 \cdot h_2 \right) \cdot \frac{\partial}{\partial s_2} \left(\sigma(s_2) \right) \cdot \frac{\partial}{\partial w_4} \left(w_3 \cdot x_3 + w_4 \cdot x_4 \right) \\
&= 2(\hat{y} - y) \cdot \sigma(s_3)(1 - \sigma(s_3)) \cdot w_6 \cdot \sigma(s_2)(1 - \sigma(s_2)) \cdot x_4 \\
&= 2(\hat{y} - y) \cdot \hat{y}(1 - \hat{y}) \cdot w_6 \cdot h_2(1 - h_2) \cdot x_4 \\
&= 2(0.586 - 0.7) \cdot 0.586(1 - 0.586) \cdot 0.5 \cdot 0.985(1 - 0.985) \cdot -2.7 \\
&\approx 0.00110
\end{aligned}
$$