CSE616: Assignment (1)

1) Suppose we have a small dense neural network as is shown in fig.1.

a) When the activation functions in the first and second layer are identity functions:

$$\begin{pmatrix} a1 \\ a2 \\ a3 \end{pmatrix} = \begin{pmatrix} w11 & w21 \\ w12 & w22 \\ w13 & w23 \end{pmatrix} \begin{pmatrix} x1 \\ x2 \end{pmatrix} + \begin{pmatrix} b1 \\ b2 \\ b3 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 1 & -1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix}$$

$$\hat{y} = (w11 \quad w21 \quad w31) \begin{pmatrix} a1 \\ a2 \\ a3 \end{pmatrix} + b1 = (3 \quad 1 \quad 2) \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix} + 1 = 36$$

b) When the activation functions in the first and second layer are ReLU functions:

$$\begin{pmatrix} a1 \\ a2 \\ a3 \end{pmatrix} = \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix} \rightarrow f(w;x) \rightarrow \begin{pmatrix} 9 \\ 0 \\ 5 \end{pmatrix}$$

$$\hat{y} = (w11 \quad w21 \quad w31) \begin{pmatrix} a1 \\ a2 \\ a3 \end{pmatrix} + b1 = (3 \quad 1 \quad 2) \begin{pmatrix} 9 \\ 0 \\ 5 \end{pmatrix} + 1 = 38$$

c) Squared loss error function $J = (\hat{y} - y)^2$, identity functions as activations in all nodes

$$\frac{\partial J}{\partial b_1} = \frac{\partial J}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial b_1} = 2(\hat{y} - y) * 1 = 2(36 - 32) = 8$$

$$\frac{\partial J}{\partial w_{21}} = \frac{\partial J}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_{21}} = 2(\hat{y} - y) \cdot a_2 = 2(\hat{y} - y) * -2 = -4(\hat{y} - y) = -4(36 - 32) = -16$$

$$\frac{\partial J}{\partial b_2} = \frac{\partial J}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial a_2} \cdot \frac{\partial a_2}{\partial b_2} = 2(\hat{y} - y) \cdot w_{21} \cdot 1 = 2(\hat{y} - y) = 2(36 - 32) = 8$$

$$\frac{\partial J}{\partial w_{13}} = \frac{\partial J}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial a_3} \cdot \frac{\partial a_3}{\partial w_{13}} = 2(\hat{y} - y) \cdot w_{31} \cdot x_1 = 2(\hat{y} - y) * 2 * 1 = 4(\hat{y} - y) = 4(36 - 32) = 16$$

d) Perform one step with gradient descent with learning rate 2

$$\frac{\partial J}{\partial b_2} = 8, \qquad \frac{\partial J}{\partial w_{13}} = 16, \qquad \eta = 2$$

$$b^{\tau+1} = b^{\tau} - \eta \nabla E(b^{\tau}),\ b_2 = 0 - 2 * 8 = 16$$

$$w^{\tau+1} = w^{\tau} - \eta E(w^{\tau}), w_{13} = 3 - 2 * 16 = 29$$

e) No, this will not be a good indicator

---

2) Compute the derivatives where $f = \sin(g_1) + g_2^2, \quad g_1 = x_1 e^{x_2}, \quad g_2 = x_1 + x_2^2$

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial g_1} \cdot \frac{\partial g_1}{\partial x_1} + \frac{\partial f}{\partial g_2} \cdot \frac{\partial g_2}{\partial x_1} = \cos(g_1)\, e^{x_2} + 2g_2$$

$$\frac{\partial f}{\partial x_2} = \frac{\partial f}{\partial g_1} \cdot \frac{\partial g_1}{\partial x_2} + \frac{\partial f}{\partial g_2} \cdot \frac{\partial g_2}{\partial x_2} = \cos(g_1)\, x_1 e^{x_2} + 4g_2 x_2$$

---

3) Compute the derivatives

1) $f(z) = \frac{1}{1 + e^{-z}} \qquad \frac{df}{dz} = \frac{e^{-z}}{(1 + e^{-z})^2}$

2) $f(\omega) = \frac{1}{1 + e^{-\omega^T x}} \qquad \frac{df}{d\omega} = \frac{x\, e^{-\omega^T x}}{\left(1 + e^{-\omega^T x}\right)^2}$

3) $J(\omega) = \frac{1}{2} \sum_{i=1}^{m} \left| \omega^T x^{(i)} - y^{(i)} \right| \qquad \frac{dJ}{d\omega} = \frac{1}{2} \sum_{i=1}^{m} \left| x^{(i)} \right|$

4) $J(\omega) = \frac{1}{2} \left[\sum_{i=1}^{m} (\omega^T x^{(i)} - y^{(i)})^2 \right] + \lambda \|\omega\|_2^2 \qquad \frac{dJ}{d\omega} = \frac{1}{2} \left[\sum_{i=1}^{m} x^{(i)} (\omega^T x^{(i)} - y^{(i)}) \right]$

5) $J(\omega) = \sum_{i=1}^{m} \left[y^{(i)} \log_a \left(\frac{1}{1 + e^{-\omega^T x^{(i)}}} \right) + (1 - y^{(i)}) \log_a \left(1 - \frac{1}{1 + e^{-\omega^T x^{(i)}}} \right) \right]$

$$\frac{dJ}{d\omega} = \sum_{i=1}^{m} \left[y^{(i)} \frac{x^{(i)}\, e^{-\omega^T x^{(i)}} (1 + e^{-\omega^T x^{(i)}})}{(1 + e^{-\omega^T x^{(i)}})^2\, lna} - (1 - y^{(i)}) \frac{x^{(i)}\, e^{-\omega^T x^{(i)}}}{\left(1 - \frac{1}{1 + e^{-\omega^T x^{(i)}}}\right) (1 + e^{-\omega^T x^{(i)}})^2\, lna} \right]$$

6) $f(\omega) = \tanh[\omega^T x] \qquad \nabla_{\omega} f = x\, \text{sech}^2[\omega^T x]$