Neural Networks and their Applications

2022 Spring

Assignment 1

2002387 Mahmoud Aboud Nada *Instructor: Dr. Hazem Abbas*

§ Home Work 1 §

Problem 1:

(1) The equation used to output ŷ for each layer is $W^T x + b$ after that we apply the identity activation function which copies the output without any change.

$$z^{[1]} = \begin{pmatrix} 2 & 2 \\ 1 & -1 \\ 3 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix}$$

$$\sigma(z^{[1]}) = \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix}$$

$$z^{[2]} = \begin{pmatrix} 3 & 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix} + 1 = 36$$

$$\sigma(z^{[2]}) = 36$$

(2) same as before but applying ReLU activation function $= \max(0, \hat{y})$

$$z^{[1]} = \begin{pmatrix} 2 & 2 \\ 1 & -1 \\ 3 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix}$$

$$\sigma(z^{[1]}) = \begin{pmatrix} 9 \\ 0 \\ 5 \end{pmatrix}$$

$$z^{[2]} = \begin{pmatrix} 3 & 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 9 \\ 0 \\ 5 \end{pmatrix} + 1 = 37$$

$$\sigma(z^{[2]}) = 37$$

(3) Using $\mathbf{j} = (\hat{\mathbf{y}} - \mathbf{y})^{\mathbf{2}}$ loss function

$$\frac{\partial j}{\partial w^{[2]}} = \frac{\partial j}{\partial a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}} \cdot \frac{\partial z^{[2]}}{\partial w^{[2]}}$$
$$\frac{\partial j}{\partial w^{[2]}} = 2 * (a^{[2]} - y) * 1 * a^{[1]}$$
$$\frac{\partial j}{\partial w^{[2]}} = 2 * (36 - 32) * 1 * \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix} = \begin{pmatrix} 72 \\ -16 \\ 40 \end{pmatrix}$$
$$\frac{\partial j}{\partial b^{[2]}} = \frac{\partial j}{\partial a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}} \cdot \frac{\partial z^{[2]}}{\partial b^{[2]}}$$
$$\frac{\partial j}{\partial b^{[2]}} = 2 * (a^{[2]} - y) * 1 * a^{[1]}$$
$$\frac{\partial j}{\partial b^{[2]}} = 2 * (36 - 32) * 1 * 1 = 8$$
$$\frac{\partial j}{\partial w^{[1]}} = \frac{\partial j}{\partial a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}} \cdot \frac{\partial z^{[2]}}{\partial a^{[1]}} \cdot \frac{\partial a^{[1]}}{\partial z^{[1]}} \cdot \frac{\partial z^{[1]}}{\partial w^{[1]}}$$
$$\frac{\partial j}{\partial w^{[1]}} = 8 * 1 * a^{[1]} * 1 * x^{[1]}$$
$$\frac{\partial j}{\partial w^{[1]}} = 8 * \begin{pmatrix} 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix} = \begin{pmatrix} 72 & -16 & 40 \\ 216 & -48 & 120 \end{pmatrix}$$
$$\frac{\partial j}{\partial b^{[1]}} = \frac{\partial j}{\partial a^{[2]}} \cdot \frac{\partial a^{[2]}}{\partial z^{[2]}} \cdot \frac{\partial z^{[2]}}{\partial a^{[1]}} \cdot \frac{\partial a^{[1]}}{\partial z^{[1]}} \cdot \frac{\partial z^{[1]}}{\partial b^{[1]}}$$
$$\frac{\partial j}{\partial b^{[1]}} = 8 * 1 * \begin{pmatrix} 9 \\ -2 \\ 5 \end{pmatrix} * 1 * 1 = \begin{pmatrix} 72 \\ -16 \\ 40 \end{pmatrix}$$
$$\frac{\partial j}{\partial b_1^{[2]}} = 8$$
$$\frac{\partial j}{\partial w_{21}^{[2]}} = -16$$
$$\frac{\partial j}{\partial b_2^{[1]}} = -16$$
$$\frac{\partial j}{\partial w_{13}^{[1]}} = 40$$

(4) using the following equations we can calculate the updated values

$$w^{[l]} = w^{(l)} - \text{learning rate} \times \frac{\partial j}{\partial w^{(l)}}$$
$$b^{[l]} = b^{(l)} - \text{learning rate} \times \frac{\partial j}{\partial b^{(l)}}$$

$$w^{[1]} = \begin{pmatrix} 2 & 1 & 3 \\ 2 & -1 & 1 \end{pmatrix} - 2 * \begin{pmatrix} 72 & -16 & 40 \\ 216 & -48 & 120 \end{pmatrix} = \begin{pmatrix} -142 & 33 & -77 \\ -430 & 95 & -239 \end{pmatrix}$$
$$b^{[1]} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - 2 * \begin{pmatrix} 72 \\ -16 \\ 40 \end{pmatrix} = \begin{pmatrix} -143 \\ 32 \\ 40 \end{pmatrix}$$
$$b_2^{[1]} = 32$$
$$w_{13}^{[1]} = -77$$

(5) splitting the data into smaller subsets is good way to either make good performance measure and tell if your model converges or not, because the small set used for training is the only thing that the model knows, using different set (test set) to measure how the model is performing can show you better if the model's performance is acceptable or not, on the other hand training model with all collected data and sticking with training set performance measure is not wise.

Problem 2:

(1)

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial g_1} * \frac{\partial g_1}{\partial x_1} + \frac{\partial f}{\partial g_2} * \frac{\partial g_2}{\partial x_1}$$
$$\frac{\partial f}{\partial x_1} = \cos(g_1 e^{x_2}) + 2g_2 * 1$$
$$\frac{\partial f}{\partial x_1} = \cos(x_1 e_2^x) * e_2^x + 2x_1 + 2x_2^2$$

(2)

$$\frac{\partial f}{\partial x_2} = \frac{\partial f}{\partial g_1} * \frac{\partial g_1}{\partial x_2} + \frac{\partial f}{\partial g_2} * \frac{\partial g_2}{\partial x_2}$$
$$\frac{\partial f}{\partial x_2} = \cos(x_1 e_2^x) + (x_1 e_2^x + e_2^x) + (2x_1 + 2x_2^2) * 2x_2$$

Problem 3:

(1)

$$f(z) = \frac{1}{1 + e^- z}$$
$$\partial \frac{(1+e^{-z})^{-1}}{\partial z} = -(1+e^{-z})^{-2} * \partial \frac{(1+e^{-z})}{\partial z}$$
$$\frac{\partial f}{\partial z} = -(1+e^{-z})^{-2} * (e^{-z} * -1)$$
$$\frac{\partial f}{\partial z} = (1+e^{-z})^{-2} * e^{-z}$$
$$\frac{\partial f}{\partial z} = \frac{e^{-z}}{(1+e^{-z})^2}$$

(2)

$$f(w) = \frac{1}{1 + e^{-w^T x}}$$
$$\partial \frac{(1+e^{-w^T x})^{-1}}{\partial w} = -(1+e^{-w^T x})^{-2} * \partial \frac{(1+e^{-w^T x})}{\partial w}$$
$$\frac{\partial f}{\partial z} = -(1+e^{-w^T x})^{-2} * (e^{-w^T x} * -x)$$
$$\frac{\partial f}{\partial w} = (1+e^{-w^T x})^{-2} * e^{-w^T x} * x$$
$$\frac{\partial f}{\partial w} = \frac{x e^{-w^T x}}{(1+e^{-w^T x})^2}$$

(3)

$$J(w) = \frac{1}{2} \sum_{i=1}^{m} |w^T x^{(i)} - y^{(i)}|$$

(4)

$$J(w) = \frac{1}{2}[\sum_{i=1}^{m}(w^T x^{(i)} - y^{(i)})^2] + \lambda||w||_2^2$$

(5)

$$J(w) = \sum_{i=1}^{m} y^{(i)} \log(\frac{1}{1+e^{-w^T x^{(i)}}}) + (1-y^{(i)}) \log(1 - \frac{1}{1+e^{-w^T x^{(i)}}})$$

(6)