Midterm Exam I- Solutions
ECE 685D– Introduction to Deep Learning
Fall 2023

Instructor: Prof. Vahid Tarokh
ECE Department, Duke University

Oct 4th 2022
10:05 AM - 11:20 AM
(Exam duration: 75 minutes)

Name: ____

Duke ID: ____

---

1. Solution:
(a) Below are the detail of the convolution between X and k_1

$$X[:,:,0] * k_1[0,:,:,0] = \begin{bmatrix} -1 & -1 & 1 & 1 \\ -3 & -3 & -1 & 2 \\ -1 & 2 & 0 & -1 \\ 1 & -2 & -4 & -4 \end{bmatrix} \tag{1}$$

$$X[:,:,1] * k_1[1,:,:,0] = \begin{bmatrix} 0 & 2 & 0 & 1 \\ 2 & 1 & 2 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{2}$$

$$X[:,:,2] * k_1[2,:,:,0] = \begin{bmatrix} 0 & 2 & 0 & 1 \\ 2 & 1 & 2 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{3}$$

Hence,

$$Y_1 = ReLU(X * k_1 + \mathbb{1}_{1\times4\times4}) = \begin{bmatrix} 0 & 2 & 2 & 3 \\ 0 & 0 & 2 & 3 \\ 1 & 3 & 1 & 0 \\ 3 & 0 & 0 & 0 \end{bmatrix} \tag{4}$$

(b) Next, the maxpooling filter is applied to the output Y_1.

$$Z_1 = maxpooling(Y_1) = \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} \tag{5}$$

$$Y_2 = ReLU(Z_1 * k_2 + 0.5 \times \mathbb{1}_{2\times2\times2}) = 0.5 \times \mathbb{1}_{2\times2\times2} \tag{6}$$

$$Z_2 = maxpooling(Y_2) = 0.5 \times \mathbb{1}_{1\times1\times2} \tag{7}$$

2. Solution:

(a)

$$\begin{aligned}\frac{\partial L}{\partial w} &= -\sum_{i=1}^{4}\frac{y_i}{p_i}\frac{\partial p_i}{\partial w} + \frac{1-y_i}{1-p_i}\frac{\partial 1-p_i}{\partial w}\\ &= -\sum_{i=1}^{4}\frac{y_i}{p_i}p_i(1-p_i)x_i - \frac{1-y_i}{1-p_i}p_i(1-p_i)x_i\\ &= -\sum_{i=1}^{4}(y_i-p_i)x_i\end{aligned}$$

$$\begin{aligned}\frac{\partial L}{\partial b} &= -\sum_{i=1}^{4}\frac{y_i}{p_i}\frac{\partial p_i}{\partial b} + \frac{1-y_i}{1-p_i}\frac{\partial 1-p_i}{\partial b}\\ &= -\sum_{i=1}^{4}\frac{y_i}{p_i}p_i(1-p_i) - \frac{1-y_i}{1-p_i}p_i(1-p_i)\\ &= -\sum_{i=1}^{4}(y_i-p_i)\end{aligned}$$

(b) $\lambda_0 = 0, \lambda_1 = 1, \lambda_2 = (1+\sqrt{1+4})/2, \lambda_3 = (1+\sqrt{1+4\lambda_2^2})/2, \gamma_1 = 0, \gamma_2 = (1-\lambda_2)/\lambda_3$

(c) We have: $\nabla L(w,b) = \left(\frac{\partial L}{\partial w}, \frac{\partial L}{\partial b}\right)$

From Nesterove Algorithm, we obtain: $t_2 = (w,b) - \nabla L(w,b)/\beta$

Hence: $w_2 = (1-\gamma_1)t_2[0] + \gamma_1 t_1[0] = t_2[0]$ and $b_2 = (1-\gamma_1)t_2[1] + \gamma_1 t_1[1] = t_2[1]$.

Similarly for t_3, w_3, b_3.

3. Solution:
   From the node graph, we can derive the output of each layer as follows:

$$h_1 = \left(x_1^{(1)}W_{1,1}^{(1)} + x_2^{(1)}W_{2,1}^{(1)}\right)^2 = (1 \times 1 + 2 \times 1)^2 = 9 \tag{8}$$

$$h_2 = \left(x_1^{(1)}W_{1,2}^{(1)} + x_2^{(1)}W_{2,2}^{(1)}\right)^2 = (1 \times 2 + 2 \times 1)^2 = 16 \tag{9}$$

   The output y is defined as follows:

$$y = \sigma(x^{(3)}) = \sigma(h_1W_1^{(2)} + h_2W_2^{(2)}) = \sigma(9 \times 0.1 + 16 \times 3) = \sigma(48.9) = \frac{1}{1 + \exp(-48.9)} \tag{10}$$

   The gradient for each weight:

$$\frac{\partial y}{\partial W_1^{(2)}} = y(1-y)h_1$$
$$\frac{\partial y}{\partial W_2^{(2)}} = y(1-y)h_2$$

$$\frac{\partial y}{\partial W_{1,1}^{(1)}} = \frac{\partial y}{\partial h_1}\frac{\partial h_1}{\partial W_{1,1}^{(1)}} = y(1-y)W_1^{(2)}2(x_1^{(1)}W_{1,1}^{(1)} + x_2^{(1)}W_{2,1}^{(1)})x_1^{(1)}$$
$$\frac{\partial y}{\partial W_{1,2}^{(1)}} = \frac{\partial y}{\partial h_2}\frac{\partial h_2}{\partial W_{1,2}^{(1)}} = y(1-y)W_2^{(2)}2(x_1^{(1)}W_{1,2}^{(1)} + x_2^{(1)}W_{2,2}^{(1)})x_1^{(1)}$$
$$\frac{\partial y}{\partial W_{2,1}^{(1)}} = \frac{\partial y}{\partial h_1}\frac{\partial h_1}{\partial W_{2,1}^{(1)}} = y(1-y)W_1^{(2)}2(x_1^{(1)}W_{1,1}^{(1)} + x_2^{(1)}W_{2,1}^{(1)})x_2^{(1)}$$
$$\frac{\partial y}{\partial W_{2,2}^{(1)}} = \frac{\partial y}{\partial h_2}\frac{\partial h_2}{\partial W_{2,2}^{(1)}} = y(1-y)W_2^{(2)}2(x_1^{(1)}W_{1,2}^{(1)} + x_2^{(1)}W_{2,2}^{(1)})x_2^{(1)}$$

   $L_{MSE} = \frac{1}{2}\sum_{i=1}^{2}(y_i - t_i)^2$