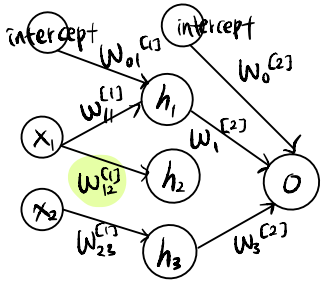
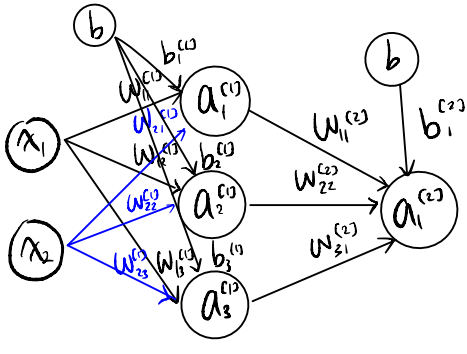
Probset #3

1.



$$loss = \frac{1}{m}\sum_{i=1}^{m}(o^{(i)} - y^{(i)})^2 \qquad y^{(i)} \in \{0,1\} \longleftarrow \text{binary classification}$$

(a) Suppose we use the sigmoid function as the activation function for h_1, h_2, h_3 and o. What is the gradient descent update to $w_{1,2}^{[1]}$, assuming we use a learning rate α? Answer in terms of $x^{(i)}$, $o^{(i)}$, $y^{(i)}$, and the weights.



$$x: \underset{(2,1)}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} \qquad W^{[1]}: \underset{(3,2)}{\begin{bmatrix} -w_1^{[1]T}- \\ -w_2^{[1]T}- \\ -w_3^{[1]T}- \end{bmatrix}} \qquad a^{[1]}: \underset{(3,1)}{\begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \\ a_3^{[1]} \end{bmatrix}} \qquad W^{[2]}: \underset{(1,3)}{[-w_1^{[2]T}-]} \qquad a^{[2]}: \underset{(1,1)}{[a_1^{[2]}]}$$

$$b^{[1]}: \underset{(3,1)}{\begin{bmatrix} b_1^{[1]} \\ b_2^{[1]} \\ b_3^{[1]} \end{bmatrix}} \qquad b^{[2]}: \underset{(1,1)}{[b_1^{[2]}]}$$

$$\mathcal{L} = (y-\hat{y})^2$$

$$a^{[2]} = g(w^{[2]}a^{[1]}+b^{[2]})$$

$$g(z) = \frac{1}{1+e^{-z}}$$

$$\frac{\partial \mathcal{L}}{\partial w^{[2]}} = \boxed{\frac{\partial \mathcal{L}}{\partial a^{[2]}} \times \frac{\partial a^{[2]}}{\partial z^{[2]}}} \times \frac{\partial z^{[2]}}{\partial w^{[2]}}$$

$$a^{[2]} = \hat{y} \in \mathbb{R}$$

$$\frac{\partial \mathcal{L}}{\partial w^{[2]}} = \frac{\partial (y-a^{[2]})^2}{\partial w^{[2]}} \quad (1,3)$$

$$= \frac{\partial}{\partial w^{[2]}} \{y - g(w^{[2]}a^{[1]}+b^{[2]})\}^2$$

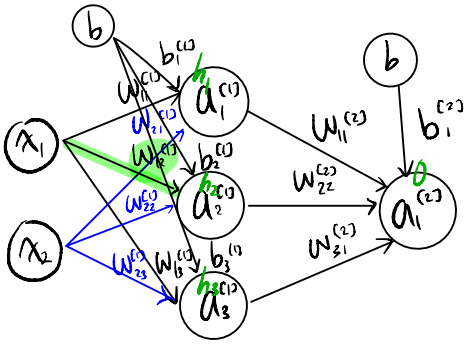
$$g(\cdots) = g(w^{[2]}a^{[1]}+b^{[2]}) = a^{[2]}$$

$$\boxed{2\{y-g(\cdots)\} \times (-1)\, g(\cdots)(1-g(\cdots))} \cdot a^{[1]T} \quad (1,3)$$

$(1,1)$ $(1,1)$

$$\frac{\partial \mathcal{L}^{(i)}}{\partial w_{12}^{[1]}} = \boxed{\frac{\partial \mathcal{L}}{\partial a^{[2]}} \times \frac{\partial a^{[2]}}{\partial z^{[2]}}} \times \underbrace{\frac{\partial z^{[2]}}{\partial a^{[1]}}}_{w^{[2]T}\ (3,1)} \times \underbrace{\frac{\partial a^{[1]}}{\partial z^{[1]}}}_{a^{[1]}\cdot(1-a^{[1]})\ (3,1)} \times \underbrace{\frac{\partial z^{[1]}}{\partial w_{12}^{[1]}}}_{x_1}$$

$(3,2)$ $(1,1)$



$$W^{[1]} = \begin{bmatrix} — & w_1^{[1]T} & — \\ — & w_2^{[1]T} & — \\ — & w_3^{[1]T} & — \end{bmatrix} = \begin{bmatrix} w_{11}^{[1]} & w_{12}^{[1]} \\ w_{12}^{[1]} & w_{22}^{[1]} \\ w_{13}^{[1]} & w_{23}^{[1]} \end{bmatrix}$$

$(3,2)$

$$w_1^{[1]} = \begin{bmatrix} w_{11}^{[1]} \\ w_{12}^{[1]} \end{bmatrix} \quad w_2^{[1]} = \begin{bmatrix} w_{12}^{[1]} \\ w_{22}^{[1]} \end{bmatrix} \quad w_3^{[1]} = \begin{bmatrix} w_{13}^{[1]} \\ w_{23}^{[1]} \end{bmatrix}$$

$$z^{[2]} = w^{[2]}a^{[1]} + b^{[2]}$$

$(1,3)$ $(3,1)$

$$z^{[1]} = w^{[1]}x + b^{[1]} = \begin{bmatrix} w_{11}^{[1]} & w_{12}^{[1]} \\ w_{12}^{[1]} & w_{22}^{[1]} \\ w_{13}^{[1]} & w_{23}^{[1]} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1^{[1]} \\ b_2^{[1]} \\ b_3^{[1]} \end{bmatrix}$$

$$= \begin{bmatrix} w_{11}^{[1]}x_1 + w_{12}^{[1]}x_2 \\ w_{12}^{[1]}x_1 + w_{22}^{[1]}x_2 \\ w_{13}^{[1]}x_1 + w_{23}^{[1]}x_2 \end{bmatrix} + b^{[1]}$$

$$= \underbrace{2\{y-a^{[2]}\} \times (-1) \times a^{[2]} \times (1-a^{[2]})}_{(1,1)} \times \underbrace{w^{[2]T} \odot a^{[1]}(1-a^{[1]})}_{(3,1)} \times \underbrace{x_1}_{(1,1)}$$

$$w_{12}^{[1]} := w_{12}^{[1]} - \alpha \times \frac{1}{m} \sum_{i=1}^{m} \left(-1\right) \left\{2(y^{(i)}-a^{[2]})\, a^{[2]}(1-a^{[2]})\, w^{[2]T} \odot a^{[1]}(1-a^{[1]})\, x_1^{(i)}\right\}$$

$$\therefore w_{12}^{[1]} := w_{12}^{[1]} + \alpha \times \frac{1}{m} \sum_{i=1}^{m} \left\{2(y^{(i)}-o^{(i)})\, o^{(i)}(1-o^{(i)})\, w_2^{[2]T} \cdot h_2(1-h_2)\, x_1^{(i)}\right\}$$

* 연결되어 있는 것만 계산 *