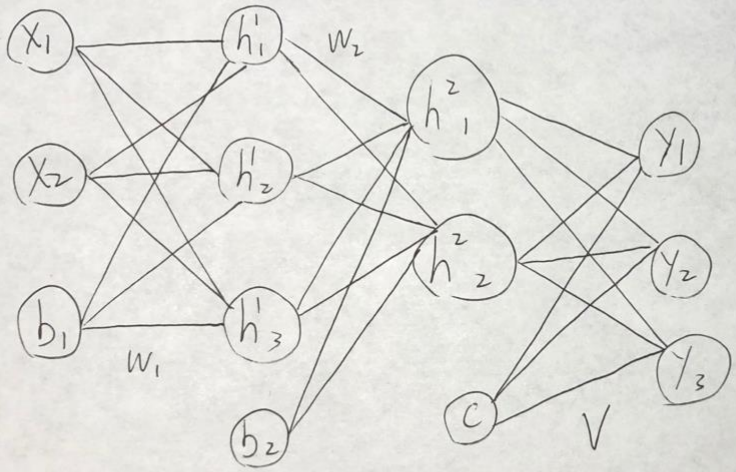
1.1



1.2

$a_1 = W^1 x + b_1$

$h_1 = \max(0, a_1)$

$a_2 = W^2 h_1 + b_2$

$h_2 = \max(0, a_2)$

$a_3 = V h_2 + c$

$\hat{y} = \text{softmax}(a_3)$

2.1

$$f(x,y) = (1-x)^2 + 100(y-x^2)^2$$

$$\frac{df}{dx} = -2(1-x) - 400x(y-x^2)$$

$$= -2 + 2x - 400xy + 400x^3$$

$$\frac{df}{dy} = 200(y-x^2) = 200y - 200x^2$$

3.1

$z_1 = W_1 x + b_1$ $\quad z_2 = W_2 h_1 + b_2$ $\quad z_3 = V h_2 + c$

$Out_1 = f(z_1)$ $\quad Out_2 = f(z_2)$ $\quad \hat{y} = softmax(z_3)$

f is ReLU, g is softmax

$$\frac{\partial L}{\partial z_3} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial a} = \sum \hat{y} - y \qquad \frac{\partial L}{\partial V} = (\hat{y} - y) \circ Out_2^T$$

$$\frac{\partial L}{\partial W_2} = [V^T(\hat{y} - y)] \circ Out_2 \cdot Out_1^T \qquad \frac{\partial L}{\partial b_2} = \sum [V^T \cdot (\hat{y} - y)] \circ Out_2$$

$$\frac{\partial L}{\partial b_1} = \sum [W_2^T \cdot V^T(\hat{y} - y) \circ Out_2] \circ Out_1$$

$$\frac{\partial L}{\partial W_1} = \{W_2^T \cdot [V^T(\hat{y} - y) \circ Out_2] \circ Out_1\} X^T$$