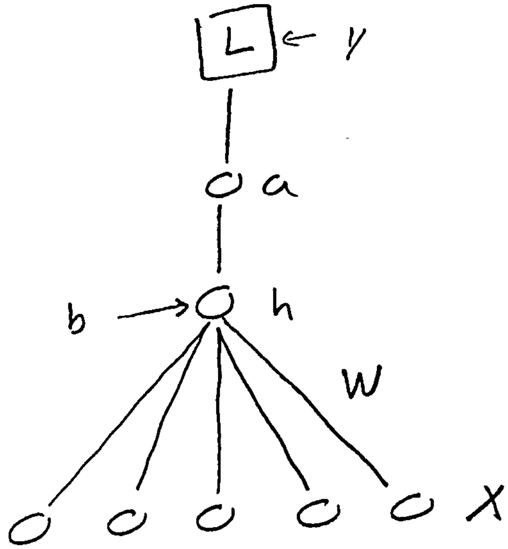
Nathan Inkawhich

Homework 1

Questions 1 + 5

1 a

Logistic Regression



$X \in \mathbb{R}^{1 \times M}$

$W \in \mathbb{R}^{M \times 1}$

$h, a, y, b \in \mathbb{R}$

$h = XW + b$

$a = \sigma(h)$

$L = -y \log a - (1-y) \log(1-a)$

Need $\frac{\partial L}{\partial W}$, $\frac{\partial L}{\partial b}$

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial h} \frac{\partial h}{\partial W} = X^T \frac{\partial L}{\partial h}$$

$$\frac{\partial L}{\partial h} = \frac{\partial L}{\partial a} \frac{\partial a}{\partial h}$$

$$\frac{\partial L}{\partial a} = -\frac{y}{a} + \frac{1-y}{1-a} = \frac{y(1-a) + a(1-y)}{a(1-a)}$$

$$= \frac{a-y}{a(1-a)}$$

$$\frac{\partial a}{\partial h} = \sigma(h)(1-\sigma(h)) = a(1-a)$$

$$\frac{\partial L}{\partial h} = \frac{a-y}{a(1-a)} \cdot a(1-a) = a-y$$

$$\boxed{\frac{\partial L}{\partial w} = \overset{\mathbb{R}^m}{X^T} \cdot \overset{\mathbb{R}}{(a-y)}} \in \mathbb{R}^M$$

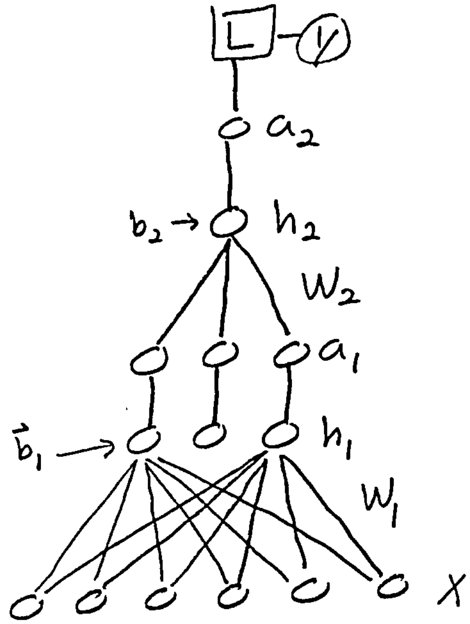
$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial h}\frac{\partial h}{\partial b}$$

$$\frac{\partial h}{\partial b} = 1$$

$$\boxed{\frac{\partial L}{\partial b} = \overset{\mathbb{R}}{a-y}} \in \mathbb{R}$$

$\boxed{1b}$

1-Hidden Layer MLP w/ binary output



$x \in \mathbb{R}^{1 \times M}$

$W_1 \in \mathbb{R}^{M \times N}$

$a_1, h_1, b_1 \in \mathbb{R}^{1 \times N}$

$W_2 \in \mathbb{R}^{N \times 1}$

$a_2, h_2, b_2 \in \mathbb{R}$

$$h_1 = x W_1 + b_1$$

$$a_1 = \sigma(h_1)$$

$$h_2 = a_1 W_2 + b_2$$

$$a_2 = \sigma(h_2)$$

$$L = -y \log a_2 - (1-y) \log(1-a_2)$$

From 1a. $\quad \dfrac{\partial L}{\partial h_2} = a_2 - y$

Need: $\dfrac{\partial L}{\partial W_1}, \dfrac{\partial L}{\partial b_1}, \dfrac{\partial L}{\partial W_2}, \dfrac{\partial L}{\partial b_2}$

$$\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial h_1}\frac{\partial h_1}{\partial W_1} = X^T \frac{\partial L}{\partial h_1}$$

$$\frac{\partial L}{\partial h_1} = \frac{\partial L}{\partial h_2}\frac{\partial h_2}{\partial a_1}\frac{\partial a_1}{\partial h_1}$$

$$\frac{\partial L}{\partial h_2} = a_2 - y \qquad \in \mathbb{R}$$

$$\frac{\partial h_2}{\partial a_1} = W_2^T$$

$$\frac{\partial a_1}{\partial h_1} = \sigma(h_1) \odot (1 - \sigma(h_1)) = a_1 \odot (1 - a_1)$$

$$\frac{\partial L}{\partial h_1} = \underbrace{(a_2 - y)}_{\mathbb{R}} \cdot \Big(\underbrace{W_2^T}_{\mathbb{R}^{1\times N}} \odot \underbrace{(a_1 \odot (1 - a_1))}_{\mathbb{R}^{1\times N}}\Big) \qquad \in \mathbb{R}^{1\times N}$$

$$\boxed{\frac{\partial L}{\partial W_1} = \underbrace{X^T}_{\mathbb{R}^{M\times 1}} \underbrace{\Big((a_2 - y) \cdot \big(W_2^T \odot (a_1 \odot (1 - a_1))\big)\Big)}_{\mathbb{R}^{1\times N}}} \qquad \in \mathbb{R}^{M\times N} \ \checkmark$$

$$\boxed{\frac{\partial L}{\partial b_1} = (a_2 - y) \cdot \big(W_2^T \odot (a_1 \odot (1 - a_1))\big)} \qquad \in \mathbb{R}^{1\times N} \ \checkmark$$

$$\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial h_2}\frac{\partial h_2}{\partial W_2} = a_1^T \frac{\partial L}{\partial h_2}$$

$$\boxed{\frac{\partial L}{\partial W_2} = \underbrace{a_1^T}_{\mathbb{R}^{N\times 1}} \cdot \underbrace{(a_2 - y)}_{\mathbb{R}}} \in \mathbb{R}^{N\times 1} \quad \checkmark$$

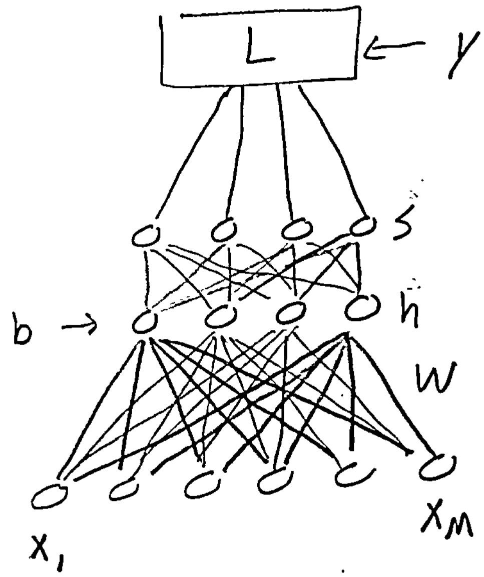
$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial h_2}\frac{\partial h_2}{\partial b_2}$$

$$\frac{\partial h_2}{\partial b_2} = 1$$

$$\boxed{\frac{\partial L}{\partial b_2} = a_2 - y} \in \mathbb{R} \quad \checkmark$$

$\boxed{1c \quad - \text{ result (a)}}$

$c = \#$ of classes



$X \in \mathbb{R}^{1 \times M}$

$W \in \mathbb{R}^{M \times c}$

$b, s, y \in \mathbb{R}^{1 \times c}$

$h = XW + b$

$$s = \text{softmax}(\vec{h_i}) = \frac{e^{h_i}}{\sum_{j=1}^{c} e^{h_j}}$$

$$L(\vec{y}, \vec{s}) = -\sum_{i=1}^{c} \mathbb{1}_{(y_i = 1)} \log(s_i) = -\vec{y} \cdot \log(\vec{s})$$

because $\vec{y}$ is one-hot

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial s} \frac{\partial s}{\partial h} \frac{\partial h}{\partial W}$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial s} \frac{\partial s}{\partial h} \frac{\partial h}{\partial b}$$

$$\frac{\partial L}{\partial \vec{s}} = \frac{\partial}{\partial \vec{s}}\left(-\vec{y} \cdot \log(\vec{s})\right) = -\frac{\vec{y}}{\vec{s}} \in \mathbb{R}^{1 \times c}$$

$$\frac{\partial \vec{s}_i}{\partial \vec{h}_j} = \frac{\partial}{\partial \vec{h}_j}\left(\frac{e^{h_i}}{\sum_{k=1}^{c} e^{h_k}}\right)$$

So two cases, when $i = j$ and when $i \neq j$

When $i = j$

$$\frac{\partial \vec{s}_i}{\partial \vec{h}_j} = \frac{\partial}{\partial h_j}\left(\frac{e^{h_i}}{\sum_K e^{h_K}}\right)$$

$$= \frac{e^{h_i}\Sigma - e^{h_i}e^{h_j}}{\Sigma^2}$$

$$= \frac{e^{h_i}(\Sigma - e^{h_j})}{\Sigma^2}$$

$$= \frac{e^{h_i}}{\Sigma} \cdot \frac{\Sigma - e^{h_j}}{\Sigma} = \boxed{s_i(1-s_j) \quad \text{when } i=j}$$

quotient rule

$$f(x) = \frac{g(x)}{h(x)}$$

$$f'(x) = \frac{g'(x)h(x) - g(x)h'(x)}{(h(x))^2}$$

let Σ represent $\sum_{K=1}^{C} e^{h_K}$

$g(x) = e^{h_i}$ $\quad h(x) = \Sigma$

When $i \neq j$

$$\frac{\partial s_i}{\partial h_j} = \frac{\partial}{\partial h_j}\left(\frac{e^{h_i}}{\Sigma}\right) = \frac{\phi\Sigma - e^{h_i}e^{h_j}}{\Sigma^2}$$

$$= -\frac{e^{h_i}}{\Sigma} \cdot \frac{e^{h_j}}{\Sigma}$$

$$= \boxed{-s_i s_j \quad \text{when } i \neq j}$$

So define Q as a $C \times C$ square matrix with $s_i(1-s_j)$ along the diagonal and $-s_i s_j$ elsewhere.

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial h}\frac{\partial h}{\partial W} = X^T \frac{\partial L}{\partial h}$$

$$\frac{\partial L}{\partial h} = \frac{\partial L}{\partial s}\frac{\partial s}{\partial h} = \underbrace{\frac{-y}{s}}_{\mathbb{R}^{1\times C}} \cdot \underbrace{Q}_{\mathbb{R}^{C\times C}} \in \mathbb{R}^{1\times C}$$

$$\boxed{\frac{\partial L}{\partial W} = \underbrace{X^T}_{\mathbb{R}^{M\times 1}} \underbrace{\left(\frac{-y}{s} Q\right)}_{\mathbb{R}^{1\times C}}} \in \mathbb{R}^{M\times C} \quad \checkmark$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial h}\frac{\partial h}{\partial b}$$

$$\frac{\partial h}{\partial b} = 1$$

$$\boxed{\frac{\partial L}{\partial b} = \underbrace{\frac{-y}{s}}_{\mathbb{R}^{1\times C}} \underbrace{Q}_{\mathbb{R}^{C\times C}}} \in \mathbb{R}^{1\times C} \quad \checkmark$$

$\boxed{\text{1c - result (b)}}$

$x \in \mathbb{R}^{1\times M}$

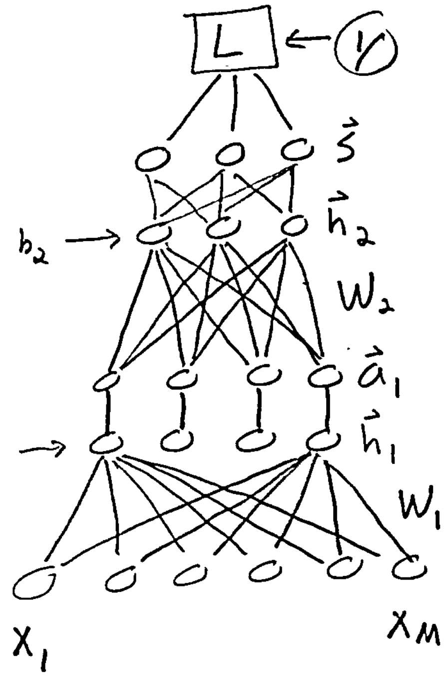
$W_1 \in \mathbb{R}^{M\times C}$

$h_1, a_1, b_1 \in \mathbb{R}^{1\times N}$

$W_2 \in \mathbb{R}^{N\times C}$

$y, h_2, b_2, s \in \mathbb{R}^{1\times C}$

$Q \in \mathbb{R}^{C\times C}$



$$\hat{h}_1 = XW_1 + b_1$$

$$a_1 = \sigma(\hat{h}_1)$$

$$\hat{h}_2 = a_1 W_2 + b_2$$

$$\vec{s} = \text{softmax}(h_2) = \frac{e^{h_{2i}}}{\sum_K e^{h_{2K}}}$$

Need $\frac{\partial L}{\partial W_2}, \frac{\partial L}{\partial b_2}, \frac{\partial L}{\partial W_1}, \frac{\partial L}{\partial b_1}$

For $\frac{\partial s}{\partial h_2}$ use same Q matrix with $s_i(1-s_i)$ on diagonal and $-s_i s_j$ elsewhere.

$$\boxed{\frac{\partial L}{\partial W_2} = \frac{\partial L}{\partial s}\frac{\partial s}{\partial h_2}\frac{\partial h_2}{\partial W_2} = a_1^T\left(\frac{-y}{s}Q\right)} = \mathbb{R}^{N\times 1}\left(\mathbb{R}^{1\times C}\mathbb{R}^{C\times C}\right) = \mathbb{R}^{N\times C} \checkmark$$

$$\boxed{\frac{\partial L}{\partial \vec{b}_2} = \frac{-y}{s}Q} \quad \mathbb{R}^{1\times C} \checkmark$$

$\mathbb{R}^{M\times 1}$ $\mathbb{R}^{1\times C}$ $\mathbb{R}^{C\times C}$ $\mathbb{R}^{C\times N}$ $\mathbb{R}^{1\times N}$

$$\boxed{\frac{\partial L}{\partial W_1} = \frac{\partial L}{\partial s}\frac{\partial s}{\partial h_2}\frac{\partial h_2}{\partial a_1}\frac{\partial a_1}{\partial h_1}\frac{\partial h_1}{\partial W_1} \Rightarrow X^T\left(\frac{-y}{s}QW_2^T \odot (a_1 \odot (1-a_1))\right)} \in \mathbb{R}^{M\times} \checkmark$$

$$\boxed{\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial s}\frac{\partial s}{\partial h_2}\frac{\partial h_2}{\partial a_1}\frac{\partial a_1}{\partial h_1}\frac{\partial h_1}{\partial b_1} \Rightarrow \frac{-y}{s}QW_2^T \odot (a_1 \odot (1-a_1))} \in \mathbb{R}^{1\times N} \checkmark$$

$\mathbb{R}^{1\times C}$ $\mathbb{R}^{C\times C}$ $\mathbb{R}^{C\times N}$ $\mathbb{R}^{1\times N}$

$\boxed{5}$

a) Yes, I was able to run on the cluster

b) This assignment took 15 - 20 hours

c) I adhered to the Duke Community standard in the completion of this assignment

NAI