C147 HW 3

1\. a. $W = U \Lambda V^T$

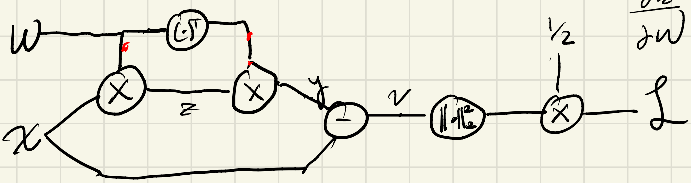
$$W^T W = V \Lambda^T \cancel{U^T U} \Lambda V^T = V \Lambda^2 V^T$$

$V \Lambda^2 V^T$ is an eigen decom

$W^T W$ is a square matrix. Minimizing $W^T W$ will minimize the eigenvalues & vectors, which in turn effects W, as its singular values are the sqrt of the eigenvalues.

b. $\mathcal{L} = \frac{1}{2} \| W^T W x - x \|_2^2$

$$\frac{\partial z}{\partial W} \frac{\partial y}{\partial z} \frac{\partial v}{\partial y} \frac{\partial \mathcal{L}}{\partial v}$$



c. $y = W^T z \qquad v = y - x \qquad \mathcal{L} = \frac{1}{2} v^T v$

$z = Wx$

$$\frac{\partial \mathcal{L}}{\partial v} = v$$

$$\frac{\partial \mathcal{L}}{\partial W} = \frac{\partial \mathcal{L}}{\partial W} + \left(\frac{\partial \mathcal{L}}{\partial W^T} \right)^T$$

$W \in m \times 1$

$x \in n \times 1$

$z \in m \times 1$

$v \in n \times 1$

d. $$\underset{m \times n}{\frac{\partial \mathcal{L}}{\partial W}} = \frac{\partial z}{\partial W} \frac{\partial y}{\partial z} \frac{\partial v}{\partial y} \frac{\partial \mathcal{L}}{\partial v} + \left(\frac{\partial y}{\partial W^T} \frac{\partial v}{\partial y} \frac{\partial \mathcal{L}}{\partial v} \right)^T$$

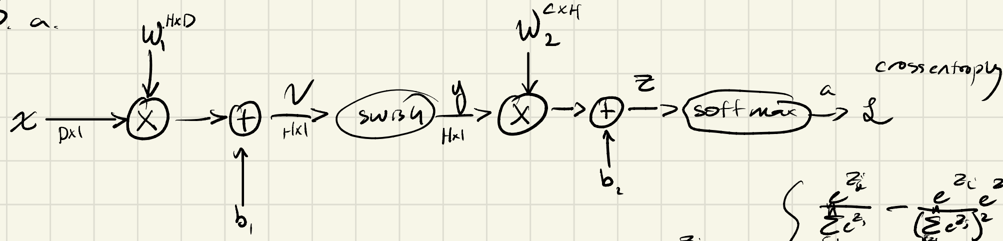
$$= (W v) x^T + (v z^T)^T$$

$$= W v x^T + z v^T$$

$$= W(W^T W x - x) x^T + Wx (W^T W x - x^T)^T$$

2\. I am a C147 student.

3. a.



b. $\mathcal{L} = \text{softmax}(z)$

$$\frac{\partial \mathcal{L}}{\partial W_2} = \frac{\partial z}{\partial W_2}\frac{\partial \mathcal{L}}{\partial z}$$

$$z = W_2 y + b_2$$

$$\nabla_{z_i} a_i = \nabla_{z_i} \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} = \begin{cases} \dfrac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} - \dfrac{e^{z_i} e^{z_i}}{\left(\sum_{j=1}^{n} e^{z_j}\right)^2} = a_i - a_i^2 = a_i(1-a_i) & i=j \\ \dfrac{e^{z_i} e^{z_j}}{\left(\sum_{j=1}^{n} e^{z_j}\right)^2} = -a_i a_j & i\neq j \end{cases}$$

$$\frac{\partial a}{\partial z} = \begin{bmatrix} a_1(1-a_1) & -a_1a_2 & \cdots & -a_1a_n \\ -a_2a_1 & \ddots & & \vdots \\ \vdots & & \ddots & -a_{n-1}a_n \\ -a_na_1 & \cdots & -a_na_{n-1} & a_n(1-a_n) \end{bmatrix} \qquad \frac{\partial \mathcal{L}}{\partial a} = \begin{bmatrix} 0 \\ \vdots \\ -\frac{1}{a_k} \\ \vdots \\ 0 \end{bmatrix}$$

$$\mathcal{L} = -\log(a_k)$$

$$\Rightarrow \frac{\partial \mathcal{L}}{\partial z} = \frac{\partial a}{\partial z}\frac{\partial \mathcal{L}}{\partial a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k - 1 \\ \vdots \\ a_n \end{bmatrix}$$

$$= \text{softmax}(z) - \mathbb{1}_{i=k}$$

$$\frac{\partial \mathcal{L}}{\partial W_2} = \left(\text{softmax}(z) - \mathbb{1}_k\right) y^T$$

$$\frac{\partial \mathcal{L}}{\partial b_2} = \text{softmax}(z) - \mathbb{1}_k$$

where

$$y = (W_1 x + b_1)\,\sigma(W_1 x + b_1)$$

$$z = W_2 y + b_2$$

c. $$\frac{\partial \mathcal{L}}{\partial W_1} = \frac{\partial v}{\partial W_1}\frac{\partial y}{\partial v}\frac{\partial \mathcal{L}}{\partial y}$$

$$\frac{\partial \mathcal{L}}{\partial y} = \frac{\partial z}{\partial y}\frac{\partial \mathcal{L}}{\partial z} = W_2\left(\text{softmax}(z) - \mathbb{1}_{i=k}\right)$$

$$y_i = v_i \sigma(v_i)$$

$$\frac{\partial y_i}{\partial v_i} = \sigma(v_i) + \sigma(v_i)(1-\sigma(v_i))$$

$$= \sigma(v_i)(2-\sigma(v_i))$$

$$\frac{\partial y}{\partial v} = \text{diag}\left(\sigma(v)(2-\sigma(v))\right)$$

$$= \text{diag}\left(\sigma(v)(2-\sigma(v))\right) W_2^T \left(\text{softmax}(z) - \mathbb{1}_{i=k}\right) x^T$$

$$\frac{\partial \mathcal{L}}{\partial b_1} = \text{diag}\left(\sigma(v)(2-\sigma(v))\right) W_2^T \left(\text{softmax}(z) - \mathbb{1}_{i=k}\right)$$