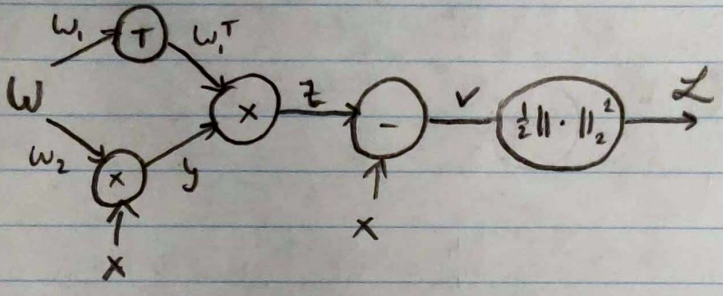
ECE 147 HW 3

Andrea Bartha

1) a. W^T is meant to be a kind of inverse operation to W. While W reduces x to a lower number of dimensions, W^T takes the m-dimensional "compressed" data and tries to restore it as well as possible. Therefore, we want the result of W^TWx to recreate the original data x as well as possible, so we try to minimize the norm of the difference multiplied by $\frac{1}{2}$ as a regularization coefficient.

b.



c. All of the derivatives for $\mathcal{L}$ which get backpropagated to W need to be summed, per the law of total derivatives, to get to the real $\frac{\partial \mathcal{L}}{\partial W}$.

d. Let us use backpropagation.

Equations: $\mathcal{L} = \frac{1}{2}\|v\|_2^2$ $\quad v = z - x$

$z = W_1^T y$ $\quad y = W_2 x$

Derivatives: $\frac{\partial \mathcal{L}}{\partial v} = \frac{1}{2}\frac{\partial}{\partial v}(v^Tv) = v$ $\quad \frac{\partial v}{\partial z} = I$

$\frac{\partial z}{\partial y} = W_1$ $\quad \frac{\partial z}{\partial W_1^T} \approx y^T$ ← per tensor trick from discussion

$\frac{\partial y}{\partial W_2} \approx x^T$ ←

Backpropagation: $\frac{\partial \mathcal{L}}{\partial z} = \frac{\partial v}{\partial z}\frac{\partial \mathcal{L}}{\partial v} = Iv = v$

$\frac{\partial \mathcal{L}}{\partial W_1^T} = \frac{\partial \mathcal{L}}{\partial z}\frac{\partial z}{\partial W_1^T} = vy^T$ ← per discussion trick

$\frac{\partial \mathcal{L}}{\partial W_1} = \left(\frac{\partial \mathcal{L}}{\partial W_1^T}\right)^T = yv^T$

$\frac{\partial \mathcal{L}}{\partial y} = \frac{\partial z}{\partial y}\frac{\partial \mathcal{L}}{\partial z} = W_1 v \cdot Wv$

$\frac{\partial \mathcal{L}}{\partial W_2} = \frac{\partial y}{\partial W_2}\frac{\partial \mathcal{L}}{\partial y} = Wvx^T$ ←

Then $\nabla_W \mathcal{L} = \frac{\partial \mathcal{L}}{\partial W} = \frac{\partial \mathcal{L}}{\partial W_1} + \frac{\partial \mathcal{L}}{\partial W_2} = yv^T + Wvx^T \cdot Wx(z-x)^T + W(z-x)$

$\nabla_W \mathcal{L} = Wx(W^TWx - x)^T + W(W^TWx - x)x^T$

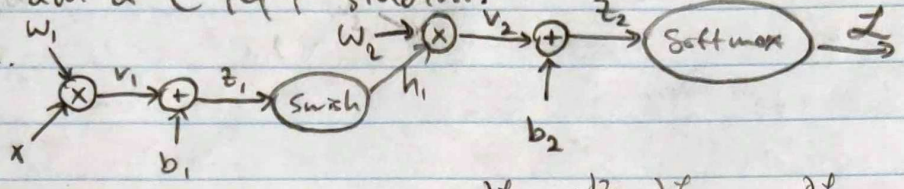
$$\nabla_W \mathcal{L} = W x((W^TW - I)x)^T + W(W^TW - I)xx^T$$
$$= W(xx^T(W^TW - I) + (W^TW - I)xx^T)$$

Since $(xx^T(W^TW - I))^T = (W^TW - I)xx^T$,

$$\nabla_W \mathcal{L} = 2W(W^TW - I)xx^T$$

2) I am a C147 student.

3) a.



(c) $z_2 = v_2 + b_2$

(c) $v_2 = W_2 h_1$

(H) $h_1 = \text{Swish}(z_1)$

(H) $z_1 = v_1 + b_1$

(H) $v_1 = W_1 x$

W_1 $H \times D$ $\quad$ W_2 $C \times H$

(c) $\frac{\partial \mathcal{L}}{\partial b_2} = \frac{\partial z_2}{\partial b_2}\frac{\partial \mathcal{L}}{\partial z_2} = \frac{\partial \mathcal{L}}{\partial z_2}$

(c) $\frac{\partial \mathcal{L}}{\partial v_2} = \frac{\partial z_2}{\partial v_2}\frac{\partial \mathcal{L}}{\partial z_2} = I\frac{\partial \mathcal{L}}{\partial z_2} = \frac{\partial \mathcal{L}}{\partial z_2}$

(H) $\frac{\partial \mathcal{L}}{\partial h_1} = \frac{\partial v_2}{\partial h_1}\frac{\partial \mathcal{L}}{\partial v_2} = W_2^T\frac{\partial \mathcal{L}}{\partial z_2}$

(C×H) $\frac{\partial \mathcal{L}}{\partial W_2} = \frac{\partial v_2}{\partial W_2}\frac{\partial \mathcal{L}}{\partial v_2} = \underset{(C\times 1)}{\frac{\partial \mathcal{L}}{\partial z_2}}\underset{(1\times H)}{h_1^T}$ ← per the discussion trick

$$\frac{\partial h_{1i}}{\partial z_{1j}} = \frac{\partial}{\partial z_{1i}}\left(\frac{z_{1i}}{1+e^{-z_{1i}}}\right)\delta_{ij}$$
$$= \frac{1+e^{-z_{1i}} + z_{1i}e^{-z_{1i}}}{(1+e^{-z_{1i}})^2}\delta_{ij}$$

$$\Rightarrow \frac{\partial h_1}{\partial z_1} = \begin{bmatrix} \frac{1+e^{-z_{11}}+z_{11}e^{-z_{11}}}{(1+e^{-z_{11}})^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1+e^{-z_{1H}}+z_{1H}e^{-z_{1H}}}{(1+e^{-z_{1H}})^2} \end{bmatrix} \quad (*)$$

(H) $\frac{\partial \mathcal{L}}{\partial z_1} = \frac{\partial h_1}{\partial z_1}\frac{\partial \mathcal{L}}{\partial h_1} = \underset{H\times H}{\frac{\partial h_1}{\partial z_1}}\underset{H\times C}{W_2^T}\underset{C\times 1}{\frac{\partial \mathcal{L}}{\partial z_2}}$

(H) $\frac{\partial \mathcal{L}}{\partial b_1} = \frac{\partial z_1}{\partial b_1}\frac{\partial \mathcal{L}}{\partial z_1} = I\frac{\partial h_1}{\partial z_1}W_2^T\frac{\partial \mathcal{L}}{\partial z_2} = \frac{\partial h_1}{\partial z_1}W_2^T\frac{\partial \mathcal{L}}{\partial z_2}$

(H) $\frac{\partial \mathcal{L}}{\partial v_1} = \frac{\partial z_1}{\partial v_1}\frac{\partial \mathcal{L}}{\partial z_1} = I\frac{\partial h_1}{\partial z_1}W_2^T\frac{\partial \mathcal{L}}{\partial z_2} = \frac{\partial h_1}{\partial z_1}W_2^T\frac{\partial \mathcal{L}}{\partial z_2}$

(H×D) $\frac{\partial \mathcal{L}}{\partial W_1} = \frac{\partial v_1}{\partial W_1}\frac{\partial \mathcal{L}}{\partial v_1} = \underbrace{\frac{\partial h_1}{\partial z_1}W_2^T\frac{\partial \mathcal{L}}{\partial z_2}}_{H\times 1}x^T$ ← discussion trick $\quad 1\times D$

Then

b) $\nabla_{W_2}\mathcal{L} = \frac{\partial \mathcal{L}}{\partial z_2}h_1^T$

$\nabla_{b_2}\mathcal{L} = \frac{\partial \mathcal{L}}{\partial z_2}$

c) $\nabla_{W_1}\mathcal{L} = \frac{\partial h_1}{\partial z_1}W_2^T\frac{\partial \mathcal{L}}{\partial z_2}x^T$

$\nabla_{b_1}\mathcal{L} = \frac{\partial h_1}{\partial z_1}W_2^T\frac{\partial \mathcal{L}}{\partial z_2}$

where $\frac{\partial h_1}{\partial z_1}$ is defined as above in (*)