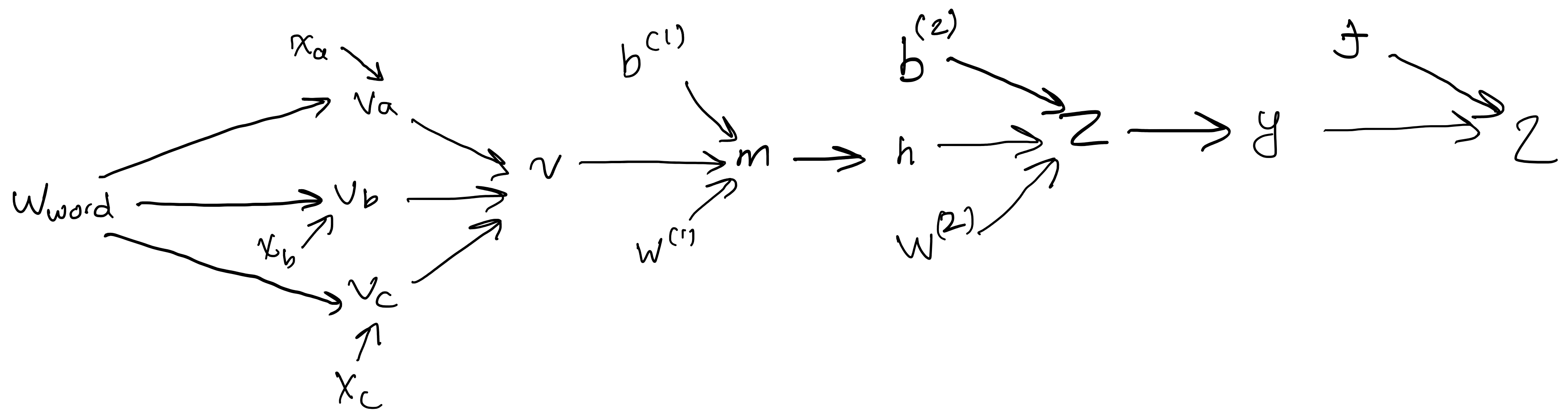
CSC413 A1 Q2.(i)

Saturday, January 29, 2022 4:31 PM

2.(i) computation graph.



→ From part (h), we justified that the update rules are identical for values with the same inputs & outputs, & who are in the same position on the graph.

→ results derived from previous questions:

update rule remains the same, just variable name changed.

$\bar{\mathcal{L}} = 1$

$\bar{y}_k = \frac{-(t+y_k) + 2y_k t}{y_k(1-y_k)}$

$$\bar{z}_i = \begin{cases} \bar{y}_i \cdot \frac{e^{z_i}(\Sigma) - e^{z_i}e^{z_i}}{(\Sigma)^2} & \text{if } i = j \\ \bar{y}_i \cdot -\frac{e^{z_j}e^{z_i}}{(\Sigma)^2} & \text{if } i \neq j \end{cases}$$

$\bar{h}_j = \bar{z}_j \cdot W^{(2)}_{ij}$ → multivariable chain rule.

$\overline{W_{ij}^{(1)}} = \bar{m}_j \cdot \cancel{x_j}$ v_j

$\overline{b_j^{(1)}} = \bar{m}_j$

$$\bar{m}_j = \begin{cases} \bar{h}_j & \text{if } m_j > 0 \\ 0 & \text{if } m_j \leq 0 \end{cases}$$

Continuing these computations:

$$\bar{v}_j = \bar{m}_j \cdot \frac{dm_j}{dv_j} \qquad \frac{dm_j}{dv_j} = \frac{d}{dv_j}\left(w_{ij}^{(1)} v_j + b_j^{(1)}\right)$$

$$= w_{ij}^{(1)}$$

→ $\bar{v}_j = \bar{m}_j \cdot w_{ij}^{(1)}$. } → multivariable

→ v_m with $m \in \{a, b, c\}$:

→ $\bar{v}_m = \bar{v}_j \cdot \frac{dv_j}{dv_{m_i}}$

$$\frac{dv_j}{dv_{m_i}} = \begin{cases} 1 & \text{if } 0 \leq j \leq 99,\ m = a \text{ or } \\ & 100 \leq j \leq 199,\ m = b \text{ or } \\ & 200 \leq j \leq 299,\ m = c \\ & \text{AND } 0 \leq i \leq 99. \\ 0 & \text{Otherwise} \end{cases}$$

→ $\overline{W^{word}_{ij}} = \bar{v}_a \cdot \frac{dv_a}{dW^{word}_{ij}} + \bar{v}_b \cdot \frac{dv_b}{dW^{word}_{ij}} + \bar{v}_c \cdot \frac{dv_c}{dW^{word}_{ij}}$

$= \bar{v}_a \cdot (x_a)_j + \bar{v}_b \cdot (x_b)_j + \bar{v}_c \cdot (x_c)_j$