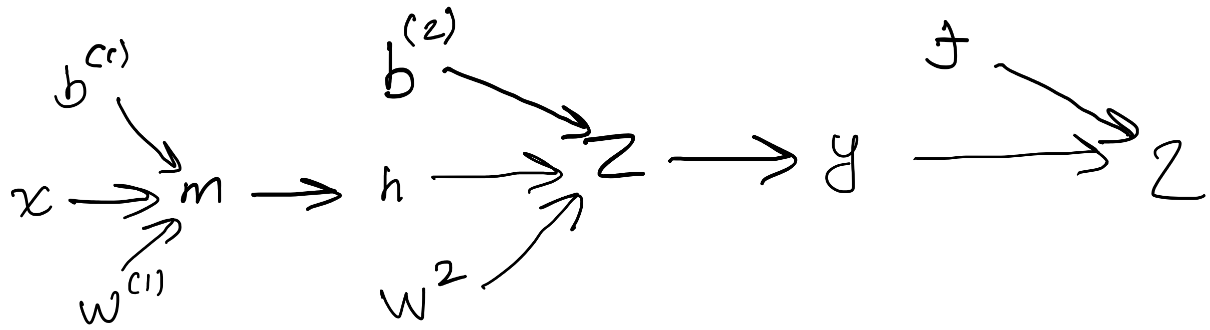
CSC413 A1 Q2.(d)

Saturday, January 29, 2022 1:14 PM

2.(d) Computation graph:



→ what we have derived from part (c):

$\bar{\mathcal{L}} = 1$ (non-vectorized).

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

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

Vectorized, this would be:

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

$\bar{y} = \frac{-(t + y) + 2yt}{y(1-y)}$.

→ division element wise.

$\bar{z} = \bar{y} \cdot (J(\text{softmax}))^T$

→ with J denoting Jacobian.

→ continuing our computations:

$$\overline{b_j^{(2)}} = \overline{z_i} \cdot \frac{\partial z_i}{\partial b_j^{(2)}}$$

$$\frac{\partial}{\partial b_j^{(2)}} \left(W_{ij}^{(2)} h_j + b_j^{(2)} \right) = 1.$$

→ $\overline{b_j^{(2)}} = \overline{z_i}$ => vectorized: $\overline{b^{(2)}} = \overline{z}$