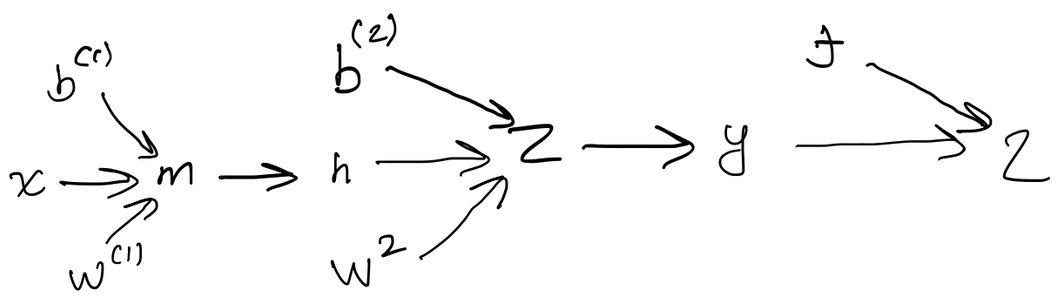
CSC413 A1 Q2.(e)

Saturday, January 29, 2022 1:25 PM

→ Find gradient descent update rule for $W^{(1)}$ and $b^{(1)}$.

Computation graph:



$$m_j = W^{(1)}_{ij} x_j + b^{(1)}_j.$$

$$z_j = W^{(2)}_{ij} h_j + b^{(2)}_j.$$

→ what we have derived from part (c):

$$\bar{L} = 1$$

$$\bar{y_k} = \frac{-(t + y_k) + 2y_k t}{y_k(1 - y_k)}. \quad \text{(non-vectorized).}$$

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

→ vectorized:

$$\bar{L} = 1$$

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

→ division element wise.

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

→ with J denoting Jacobian.

→ continuing our computations:

$$\bar{h_j} = \bar{z_j} \cdot \frac{dz_j}{dh_j} = W^{(2)}_{ij}. \Rightarrow$$

$$\bar{h_j} = \bar{z_j} \cdot W^{(2)}_{ij}.$$

$$\bar{m_j} = \bar{h_j} \cdot \frac{dh_j}{dm_j} = \frac{d}{dm_j}(\text{ReLU}(m_j))$$

$$= \begin{cases} 1 & \text{if } m_j > 0 \\ 0 & \text{if } m_j \leq 0. \end{cases}$$

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

$$\bar{W^{(1)}_{ij}} = \bar{m_j} \cdot \frac{\partial m_j}{\partial W^{(1)}_{ij}} = x_j$$

$$= \bar{m_j} \cdot x_j$$

$$\bar{b^{(1)}_j} = \bar{m_j} \cdot \frac{\partial m_j}{\partial b^{(1)}_j} = 1$$

$$= \bar{m_j}$$

→ non-vectorized, we have the following update rules:

$$\bar{h_j} = \bar{z_j} \cdot W^{(2)}_{ij}$$

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

$$\overline{W^{(1)}_{ij}} = \bar{m_j} \cdot x_j$$

$$\overline{b^{(1)}_j} = \bar{m_j}$$

⇒ vectorized:

$$\bar{h} = W^{(2)} \cdot (\bar{z})^T$$

$$\bar{m} = \max(\bar{h}, 0).$$

$$\overline{W^{(1)}} = \bar{m} \cdot X^T$$

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