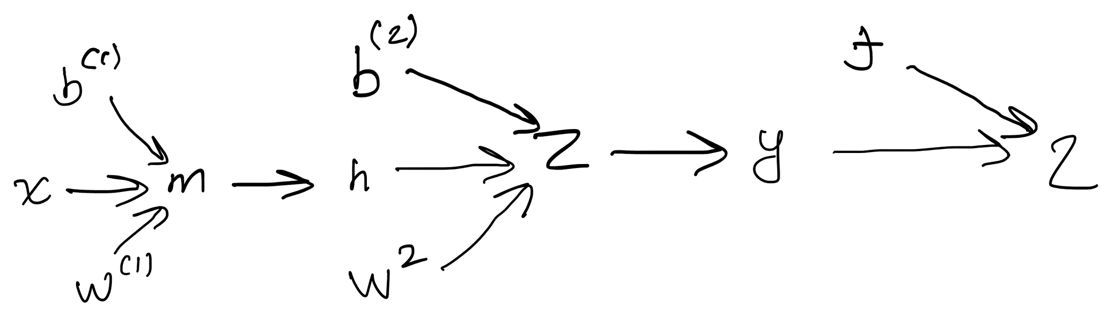
CSC311 A1 Q2.(b) + Q2.(c)

Saturday, January 22, 2022 6:58 PM

2.(b) Computation graph:



2.(c) $\bar{\mathcal{L}} = 1$

→ Computing update rule for $W^{(2)}_{ij}$.

→ $z_i = W^{(2)}_{ij} h_j + b^{(2)}_j$

$\bar{y}_k = \bar{\mathcal{L}} \cdot \frac{d\mathcal{L}}{dy_k}$

$= \frac{d}{dy_k}\left(-t\log(y_k) - (1-t)\log(1-y_k)\right)$

$= -t \cdot \frac{1}{y_k} - (1-t)\cdot\frac{1}{1-y_k}$.

$= -\frac{t}{y_k} - \frac{(1-t)}{1-y_k} = \frac{-t(1-y_k) - (1-t)y_k}{y_k(1-y_k)}$

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

$f\left(\begin{bmatrix}2\\3\\5\end{bmatrix}\right) = \begin{pmatrix}0.03\\0.06\\0.91\end{pmatrix}$

$\frac{dS_i}{da_j}$ = output w.r.t. ith output → jth input.

$\bar{z}_i = \bar{y}_k \cdot \frac{dy_k}{dz_i}$

$\frac{d}{dz_i}(\text{softmax}(z_i))$

$= \frac{d}{dz_i}\left(\frac{e^{z_i}}{\sum_i e^{z_m}}\right) = \frac{d}{dz_i}\left(\frac{e^{z_i}}{e^{z_1}+\dots+e^{z_m}}\right)$

= computing Jacobian matrix for softmax function:

Let i denote the ith output & j denote the jth input.

if $i = j$: let $\Sigma_i = e^{z_i} + \dots + e^{z_m}$

$\frac{d}{dz_i}\left(\frac{e^{z_i}}{e^{z_i}+\dots+e^{z_m}}\right)$

⇒ By quotient rule:

$\frac{e^{z_i}(\Sigma_i) - e^{z_j}e^{z_i}}{(\Sigma_i)^2}$.

if $i \neq j$:

again, by quotient rule:

$\frac{0(\Sigma_i) - e^{z_j}e^{z_i}}{(\Sigma_i)^2} = \frac{-e^{z_j}e^{z_i}}{(\Sigma_i)^2}$.

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

$\overline{W_{ij}^{(2)}} = \bar{z}_i \cdot \frac{dz_i}{dw_{ij}}$

$\frac{d}{w_{ij}}\left(w_{ij}^{(2)} h_j + b_j\right)$

$= h_j$

$= \bar{z}_i \cdot h_j$

Overall, we have that non-vectorized, the update rule for $W_{ij}^{(2)}$ is:

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

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

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

$\overline{W_{ij}^{(2)}} = \bar{z}_i \cdot h_j$

or

$W_{ij}^{(2)} \leftarrow W_{ij}^{(2)} - \alpha\, \overline{W_{ij}^{(2)}}$

Vectorized, this would be:

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

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

⇒ division element wise.

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

→ with J denoting Jacobian.

$\boxed{\overline{W^{(2)}} = \bar{z}\, h^T}$