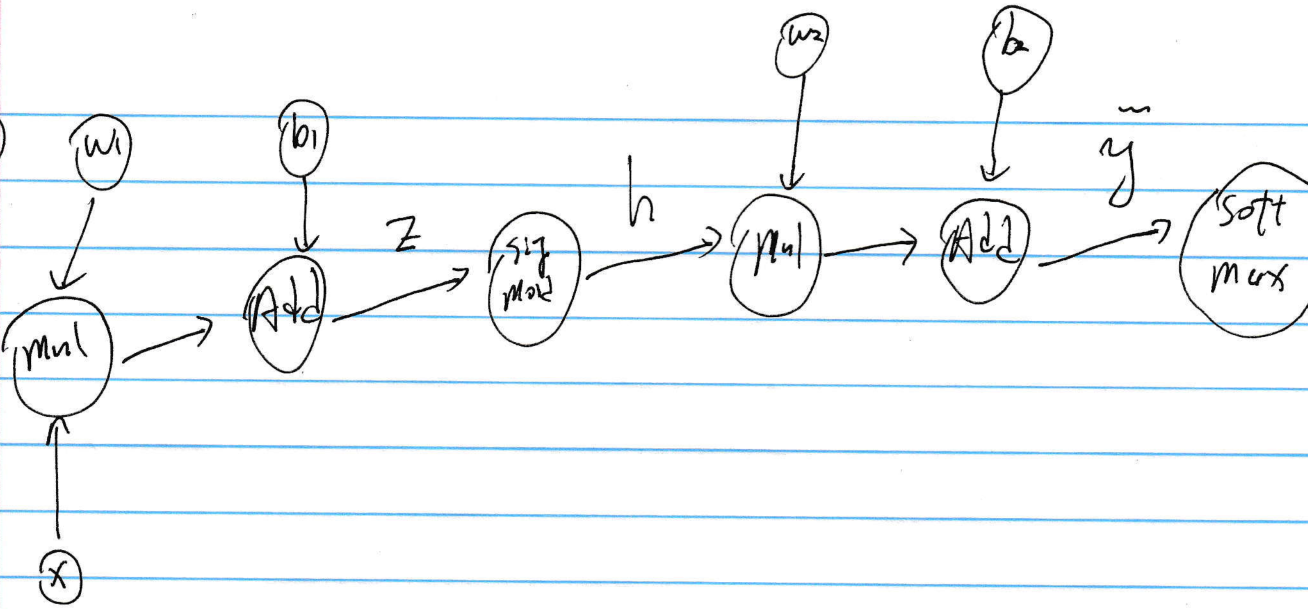
1. (a)



(b) $d(\hat{y}_i) = \frac{-y_i}{\hat{y}_i} = (-y \log \hat{y}_i)'$

(c) $$\frac{\partial \hat{y}_i}{\partial \tilde{y}_j} = \frac{e^{\tilde{y}_i}\left(\sum_n e^{\tilde{y}_n} - e^{\tilde{y}_j}\right)}{\left(\sum e^{\tilde{y}_n}\right)^2}$$ when $i = j$

$= P_i(1-P_j)$ assume $P_i = \frac{e^{\tilde{y}_i}}{\sum e^{\tilde{y}_n}} = \hat{y}$

$= \hat{y}_i(1-\hat{y}_j)$

when $i \neq j$

$$= \frac{-e^{\tilde{y}_i + \tilde{y}_j}}{\sum e^{\tilde{y}_k}} = -P_i P_j = \hat{y}_i \hat{y}_j$$

$$\frac{\partial \hat{y}}{\partial \tilde{y}} = P_i(\delta_{ij} - P_j)$$ where $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$

$= \hat{y}_i(\delta_{ij} - \hat{y}_j)$

chain rule

$$\cancel{d\hat{y}_i}$$

$$d\tilde{y}_i = \frac{\partial J}{\partial \tilde{y}} = -\sum_k y_k \frac{\partial \log(\hat{y})}{\partial \tilde{y}} = -\sum y_k \frac{1}{\hat{y}_k} * \frac{\partial \hat{y}_k}{\partial \tilde{y}_i}$$

$$= -y_i(1-p_i) + \sum_{k \neq i} p_i * y_k = -y_i + p_i(\underbrace{y_i + \sum_{k\neq i} y_k}_{=1})$$

$$d(\tilde{y}) = p_i - y_i = \hat{y}_i - y_i$$

On the right hand-side: $\frac{-y_i}{\hat{y}_i} * \hat{y}_i - \sum_j \frac{-y_j}{\hat{y}_j} \hat{y}_j \hat{y}_i$

$$= -y_i + \sum_j y_j \hat{y}_i = \boxed{\hat{y}_i - y_i} = d(\tilde{y})$$

(d) $d(h)_i = \sum_j \frac{\partial (\tilde{y})_j}{\partial h_i} d(\tilde{y})_j = (\hat{y}_i - y_i) W^{2^T}$

(e) $d(z)_i = \frac{\partial J}{\partial h_i} * \frac{\partial h}{\partial z} = \frac{e^{-z_i}}{(1+e^{-z_i})^2} * (\hat{y}_i - y_i) W^{2^T}$

$$= \sigma(z_i)(1-\sigma(z_i)) * (\hat{y}_i - y_i) W^{2^T}$$

(f) $d(W_1)_{ij} = \frac{\partial J}{\partial z_i} \frac{\partial z_i}{\partial w_{ij}} = X^T (\hat{y}_i - y_i) W^{2^T} * h_i(1-h_i)$