ML

HW5

Q1)

a) $$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{3} e^{z_j}}$$

$z_1 = 3,\ z_2 = 1,\ z_3 = -2$

$e^{z_1} = e^3 = 20.0855$
$e^{z_2} = e^1 = 2.7183$
$e^{z_3} = e^{-2} = 0.1353$

$$\sum_{j=1}^{3} e^{z_j} = 20.0855 + 2.7183 + 0.1353$$
$$= 22.9391$$

$$\sigma(z_1) = \frac{20.0855}{22.9391} \simeq 0.8759$$

$$\sigma(z_2) = \frac{2.7183}{22.9391} = 0.1185$$

$$\sigma(z_3) = \frac{0.1353}{22.9391} = 0.0059$$

b) $z_i' = z_i + c$

$$\sigma(z_i') = \frac{e^{z_i + c}}{\sum_j e^{z_j + c}} = \frac{e^{z_i} \cancel{e^c}}{\sum_j e^{z_j} \cancel{e^c}} = \frac{e^{z_i}}{\sum_j e^{z_j}} = \sigma(z)$$

$$\therefore \sigma(z_i + c) = \sigma(z_i)$$

c) $\sigma(z_i) = \dfrac{e^{z_i}}{\sum_j e^{z_j}}$

$S = \sum_j e^{z_j}$ $\qquad \sigma(z_i) = \dfrac{e^{z_i}}{S}$ —— 1

$$\frac{\partial \sigma(z_i)}{\partial z_j} = \frac{\frac{\partial}{\partial z_j} e^{z_i} \cdot S - e^{z_i} \cdot \frac{\partial S}{\partial z_j}}{S^2}$$

$$\frac{\partial e^{z_i}}{\partial z_j} = \begin{cases} e^{z_j}, & i = j \\ 0, & i \neq j \end{cases} = e^{z_i} \delta_{ij}$$

$$\frac{\partial S}{\partial z_j} = \frac{\partial}{\partial z_j} \sum_k e^{z_k} = e^{z_j}$$

$$\frac{\partial \sigma(z_i)}{\partial z_j} = \frac{e^{z_i} \delta_{ij} \cdot S - e^{z_i} \cdot e^{z_j}}{S^2}$$

from 1

$$\frac{\partial \sigma(z_i)}{\partial z_j} = \sigma(z_i)(\delta_{ij} - \sigma(z_j)) \begin{cases} \sigma(z_i)(1 - \sigma(z_i)), & i = j \\ -\sigma(z_i)\sigma(z_j), & i \neq j \end{cases}$$

d)

(i) $(z_1, z_2, z_3) = (5, 2, -1)$

$$e^{z_1} = e^5 = 148.413$$
$$e^{z_2} = e^2 = 7.389$$
$$e^{z_3} = e^{-1} = 0.3679$$

$$\sum_{j=1}^{3} e_j = 148.413 + 7.389 + 0.3679$$
$$= 156.17$$

$$\sigma(z_1) = \frac{148.413}{156.17} = 0.9503$$

$$\sigma(z_2) = \frac{7.389}{156.17} = 0.0473$$

$$\sigma(z_3) = \frac{0.3679}{156.17} = 0.0024$$

(ii) $(z_1, z_2, z_3) = (50, 20, -10)$

$$e^{z_1} = e^{50}, \quad e^{z_2} = e^{20}, \quad e^{z_3} = e^{-10}$$

As e^{50} is very large, there will be numerical overflow issues

To solve this numerical instability, substract the maximum logit before softmax

$$\boxed{\tilde{z}_i = z_i - \max_j z_j}$$

c) Taking the ratio

$$\frac{\sigma(z_1)}{\sigma(z_2)} = \frac{\dfrac{e^{z_1}}{\sum_j e^{z_j}}}{\dfrac{e^{z_2}}{\sum_j e^{z_j}}} = \frac{e^{z_1}}{e^{z_2}} = e^{z_1} - e^{z_2}$$

when $z_1 - z_2$ increases, the ratio increases exponentially, making the larger logits even more dominant in the probability distribution.