ML HW3

Q1) a) $\sigma(u) = \dfrac{1}{1+e^{-u}}$

S.T $\sigma(u)$ & $\sigma'(u)$ satisfies

1) $1-\sigma(u) = e^{-u}\sigma(u)$

$$1-\sigma(u) = 1 - \frac{1}{1+e^{-u}}$$

$$= \frac{\cancel{1}+e^{-u}-\cancel{1}}{1+e^{-u}}$$

$$= \frac{e^{-u}}{1+e^{-u}} = e^{-u}\sigma(u)$$

2) $\sigma'(u) = e^{-u}\sigma(u)^2 = \sigma(u)(1-\sigma(u))$

$$\sigma'(u) = \frac{d}{du}\left(\frac{1}{1+e^{-u}}\right)$$

$$= \frac{0(1+e^{-u}) + 1(-e^{-u})}{(1+e^{-u})^2}$$

$$= \frac{e^{-u}}{(1+e^{-u})^2} = e^{-u}\sigma(u)^2$$

from 1) we know $1-\sigma(u) = e^{-u}\sigma(u)$

$$\Rightarrow \sigma'(u) = (1-\sigma(u))(\sigma(u))$$

b) 1) S.T $\dfrac{dJ}{db} = \dfrac{1}{m} \sum_{i=1}^{m} \left(\sigma(\underline{w} \cdot \underline{x}^{(i)} + b) - y^{(i)} \right)$

Cost funct^n : $J(w,b) = \dfrac{1}{m} \sum_{i=1}^{m} \ell^{(i)}$

$$\ell^{(i)} = -y^{(i)} \log(\sigma(z)) - (1-y^{(i)}) \log(1-\sigma(z))$$

where, $z = \underline{w} \cdot \underline{x}^{(i)} + b$ & $\sigma(z) = \dfrac{1}{1+e^{-z}}$

1) gradient wrt b

$$\frac{dJ}{db} = \frac{1}{m} \sum_{i=1}^{m} \frac{d}{db} \ell^{(i)}$$

$$\frac{d}{db} \ell^{(i)} = \frac{d}{dz} \ell^{(i)} \cdot \frac{dz}{db}$$

$$\frac{dz}{db} = 1$$

$$\frac{d}{dz} \ell^{(i)} = -y^{(i)} \frac{1}{\sigma(z)} \sigma'(z) + (1-y^{(i)}) \frac{1}{1-\sigma(z)} (-\sigma'(z))$$

$$\sigma'(z) = \sigma(z)(1-\sigma(z))$$

$$= -y^{(i)} \frac{1}{\cancel{\sigma(z)}} \cancel{\sigma(z)}(1-\sigma(z)) + (1-y^{(i)}) \frac{1}{\cancel{1-\sigma(z)}} (-\sigma(z)\cancel{(1-\sigma(z))})$$

$$= -y^{(i)}(1-\sigma(z)) + (1-y^{(i)})(-\sigma(z))$$

$$\frac{d}{dz} \ell^{(i)} = \sigma(z) - y^{(i)}$$

$$\therefore \frac{dJ}{db} = \frac{1}{m} \sum_{i=1}^{m} (\sigma(\cancel{z}) - y^{(i)})$$

$$\boxed{\frac{dJ}{db} = \frac{1}{m} \sum_{i=1}^{m} \left(\sigma(\underline{w} \cdot \underline{x}^{(i)} + b) - y^{(i)} \right)}$$

2) gradient wrt w_j

$$\frac{dJ}{dw_j} = \frac{1}{m} \sum_{i=1}^{m} \frac{d}{dw_j} \ell^{(i)}$$

$$\frac{d}{dw_j} \ell^{(i)} = \frac{d}{dz} \ell^{(i)} \cdot \frac{dz}{dw_j}$$

$$\frac{d}{dz} \ell^{(i)} = \sigma(z) - y^{(i)}$$ —— from above part

since, $\frac{dz}{dw_j} = \underline{x}_j^{(i)}$

$$\frac{d}{dw_j} \ell^{(i)} = (\sigma(z) - y^{(i)})\, \underline{x}_j^{(i)}$$

$$\Rightarrow \boxed{\frac{dJ}{dw_j} = \frac{1}{m} \sum_{i=1}^{m} \underline{x}_j^{(i)} \left(\sigma(\underline{w} \cdot \underline{x}^{(i)} + b) - y^{(i)} \right)}$$