Q ① $P(y=1|x,w) = \sigma(w^Tx) = \dfrac{1}{1+e^{-w^Tx}}$

$$\frac{d\sigma(w^Tx)}{dx} = \frac{d\,(1+e^{-w^Tx})^{-1}}{dx}$$

$$= (-1)(1+e^{-w^Tx})(-e^{-w^Tx})$$

$$= \left(\frac{(1+e^{-w^Tx})-1}{1+e^{-w^Tx}}\right)\left(\frac{1}{1+e^{-w^Tx}}\right)$$

$$= \left(1 - \frac{1}{1+e^{-w^Tx}}\right) \cdot \frac{1}{1+e^{-w^Tx}}$$

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

$$= \boxed{(1-\sigma(w^Tx))(\sigma(w^Tx))} \quad — ①$$

Q2. $P(y=1|x,w) = \sigma(w^Tx) = \dfrac{1}{1+e^{-w^Tx}}$

$P(y=-1|x,w) = 1-\sigma(w^Tx) = 1 - \dfrac{1}{1+e^{-w^Tx}}$

$$= \frac{e^{-w^Tx}}{(1+e^{-w^Tx})} = \frac{1}{1+e^{w^Tx}}$$

$$\Rightarrow \boxed{P(y=\pm1|x,w) = \frac{1}{1+e^{-yw^Tx}}} \quad — ②$$

$$\text{likelihood} = \prod_{i=1}^{N} P(y_i / x_i, w)$$

from ② we can replace $P(y_i/x_i, w) = \dfrac{1}{1+e^{-y_i w^T x_i}}$

likelihood) $$L(w) = \prod_{i=1}^{N} \frac{1}{1+e^{-y_i w^T x_i}}$$

optimal weight vector for w^* can be obtained by maximizing $L(w) \Rightarrow w^* = \arg\max_w L(w)$

or we can minimize $-ve\ L(w)$

$$\Rightarrow w^* = \arg\max_w (-L(w))$$

$$E(w) = \prod_{i=1}^{N} \frac{1}{1+e^{-y_i w^T x_i}}$$

taking Log to convert $\prod \longrightarrow \sum$

$$LL(w) = \sum_{i=1}^{N} \ln\left(-\frac{1}{1+e^{-y_i w^T x_i}}\right)$$

$$= \sum_{i=1}^{N} \cancel{\ln}\left(\cancel{}(1+e^{-y_i w^T x_i}\right)$$

$$\boxed{LL(w) = \sum_{i=1}^{N} \ln\left(1+e^{-y_i w^T x_i}\right)}$$

Hence Proved.