1.) $X = \{(x_i, y_i) \ldots (x_n, y_n)\}$ $\quad x_i \in \mathbb{R}^d$ $\quad y_i \in \{0, 1\}$

$a \in \mathbb{R}$ $\quad \sigma(a) = \frac{1}{1+e^{-a}}$

$$L(y_i|x_i; w) = y_i \log(\sigma(w^T x_i)) + (1-y_i)\log(\sigma(-w^T x_i))$$

(1) Make arg to σ the same to simplify things

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

$$= \frac{e^{-a}}{e^{-a}+1} \quad \text{(multiply top and bottom by } e^{-a}\text{)}$$

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

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

$$\therefore \sigma(-a) = \sigma(a)$$

let $a_i = w^T x_i$

$$L(y_i|x_i; w) = y_i \log(\sigma(a_i)) + (1-y_i)\log(1-\sigma(a_i))$$

(2) use chain rule

let $z_i = \sigma(a_i)$

$$L(y_i|x_i; w) = y_i \log(z_i) + (1-y_i)\log(1-z_i)$$

$$\frac{\partial L(y_i|x_i; w)}{\partial w} = \frac{\partial L}{\partial z_i}\frac{\partial z_i}{\partial a_i}\frac{\partial a_i}{\partial w}$$

y_i doesn't depend on w

$$L(y_i|x_i,w) = y_i \log(z_i) + (1-y_i)\log(1-z_i)$$

(3) Find $\frac{\partial L}{\partial z_i}$

$$\frac{\partial L}{\partial z_i} = \frac{y_i}{z_i} + -\left(\frac{(1-y_i)}{(1-z_i)}\right)$$

chain rule

log derivative rule

$$= \frac{y_i(1-z_i) - z_i(1-y_i)}{z_i(1-z_i)}$$

$$= \frac{y_i - z_i}{z_i(1-z_i)}$$

(4) Find $\frac{\partial z_i}{\partial a_i}$ $\quad z_i = \sigma(a_i)$

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

$$\frac{\partial \sigma(x)}{\partial x} = -(1+e^{-x})^{-2} \cdot -e^{-x} = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \cancel{\frac{e^{-x}}{1+2e^{-x}+e^{-2x}}} \; \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{(1+e^{-x})}$$

... from previous result

$$= \cancel{\sigma(x)(1-x)} \; \sigma(x)\,(1-\sigma(x))$$

$$\frac{\partial z_i}{\partial a_i} = z_i(1-z_i)$$

$$a_i = w^T x_i$$

(5) Find $\frac{\partial a_i}{\partial w} = x_i$ (power rule / constant rule)

(6) Put it together (Yay!)

$$\frac{\partial L(y_i|x_i,w)}{\partial w} = \frac{\partial L}{\partial z_i}\frac{\partial z_i}{\partial a_i}\frac{\partial a_i}{\partial w}$$

$$= \frac{y_i - z_i}{z_i(1-z_i)} \cdot z_i(1-z_i) \cdot x_i$$

$$\frac{\partial L}{\partial w} = (y_i - z_i) \cdot x_i$$

$$= (y_i - \sigma(w^T x_i)) \cdot x_i$$

$$\cancel{\frac{\partial L(y_i|x_i,w)}{\partial w_j} = (y - \sigma(w^T x_i))[x_i]_j}$$

$$\boxed{\frac{\partial L(y_i|x_i,w)}{\partial w_j} = (y_i - \sigma(w^T x_i))\, x_{ij}}$$

$$\frac{\partial}{\partial w}(1-y_i(w^T x_i+b)) = -y_i x_i$$

$$\frac{\partial}{\partial w}(0) = 0$$

3.) $X = \{(x_i, y_i) \ldots (x_n, y_n)\}$ $\quad x_i \in \mathbb{R}^d$

$y_i \in \{-1, 1\}$

$$f(w) = \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n} \max(0, 1-y_i(w^T x_i + b))$$

$$\frac{\partial f}{\partial w} = w + C \cdot \sum_{i=1}^{n}\left([1-y_i(w^T x_i + b) > 0] \cdot -y_i x_i\right)$$

where $[x] = \begin{cases} 0 & \text{if } x \text{ false} \\ 1 & \text{if } x \text{ true} \end{cases}$

iverson bracket

$$\boxed{\frac{\partial f(w)}{\partial w_j} = w_j + C\sum_{i=1}^{n}\left([1-y_i(w^T x_i+b) > 0] \cdot -y_i x_{ij}\right)}$$