1. $L(W) = \sum_l Y^l \ln P(Y^l = 1 | X^l, W) + (1 - Y^l) \ln P(Y^l = 0 | X^l, W)$.

$\Leftrightarrow \sum_l Y^l \ln P(Y^l = 1 | X^l, W) - Y^l \ln P(Y^l = 0 | X^l, W) + \ln P(Y^l = 0 | X^l, W)$.

$\Leftrightarrow \sum_l Y^l \ln \dfrac{P(Y^l = 1 | X^l, W)}{P(Y^l = 0 | X^l, W)} + \ln P(Y^l = 0 | X^l, W)$.

$$P(Y^l = 1 | X^l, W) = \frac{\exp(W_0 + \sum_i^n W_i X_i^l)}{1 + \exp(W_0 + \sum_i^n W_i X_i^l)};$$

$$P(Y^l = 0 | X^l, W) = \frac{1}{1 + \exp(W_0 + \sum_i^n W_i X_i^l)};$$

sehingga $\dfrac{P(Y^l = 1 | X^l, W)}{P(Y^l = 0 | X^l, W)} = \exp(W_0 + \sum_i^n W_i X_i^l)$.

$\Leftrightarrow \sum_l Y^l \ln \exp(W_0 + \sum_i^n W_i X_i^l) - \ln(1 + \exp(W_0 + \sum_i^n W_i X_i^l))$.

$$\frac{dL}{dW_i} = \frac{d}{dW_i} \sum_l Y^l \ln \exp(W_0 + \sum_i^n W_i X_i^l) - \ln(1 + \exp(W_0 + \sum_i^n W_i X_i^l))$$

$$= \sum_l Y^l \frac{d \ln \exp(W_0 + \sum_i^n W_i X_i^l)}{dW_i} - \frac{d}{dW_i} \ln(1 + \exp(W_0 + \sum_i^n W_i X_i^l))$$

$$= \sum_l Y^l \frac{d}{dW_i}(W_0 + \sum_i^n W_i X_i^l) - \frac{1}{1 + \exp(W_0 + \sum_i^n W_i X_i^l)} \cdot \frac{d}{dW_i}(1 + \exp(W_0 + \sum_i^n W_i X_i^l))$$

$$= \sum_l Y^l \sum_i^n X_i^l - \frac{\exp(W_0 + \sum_i^n W_i X_i^l)}{1 + \exp(W_0 + \sum_i^n W_i X_i^l)} \cdot \frac{d}{dW_i}(W_0 + \sum_i^n W_i X_i^l)$$

$$= \sum_l Y^l \sum_i^n X_i^l - P(Y^l = 1 | X^l, W) \cdot \sum_i^n X_i^l$$

$$= \sum_l \sum_i^n X_i^l (Y^l - P(Y^l = 1 | X^l, W)).$$

2. a. parameter yang diperlukan adalah bobot w untuk tiap kelas m. Jumlah bobot yang dibutuhkan ~~adalah~~ untuk tiap vektor w_i adalah n buah. ~~untuk~~ Sehingga untuk sejumlah kelas m dibutuhkan w sejumlah $n \cdot m$

b. $$\sum_{i=1}^{s} \ln(P(Y = y_i \mid X = x_i) = \sum_{i=1}^{K} \ln \frac{\exp(w_{k0} + \sum_{i=1}^{n} w_{ki} x_i)}{1+\sum_{m}^{k-1} \exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)}$$

$$\Leftrightarrow \sum_{i=1}^{K} \ln \exp(w_{k0} + \sum_{i=1}^{n} w_{ki} x_i) - \ln(1+\sum_{m}^{k-1} \exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)).$$

$$= \sum_{i=1}^{k} w_{k0} + \sum_{i=1}^{n} w_{ki} x_i - \ln(1+\sum_{m}^{k-1} \exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)).$$

c. gradien $L = \frac{dL}{dw_i} = \frac{d}{dw_i} \sum_{i=1}^{k} w_{k0} + \sum_{i=1}^{n} w_{ki} x_i - \ln(1+\sum_{m}^{k-1} \exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i))$

$$= \sum_{i=1}^{k} \sum_{i=1}^{n} x_i - \frac{d}{dw_i} \ln(1+\sum_{m}^{k-1} \exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)).$$

$$= \sum_{i=1}^{k} \sum_{i=1}^{n} x_i - \frac{1}{1+\sum_{m}^{k-1} \exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)} \cdot \frac{d}{dw_i} (1+\sum_{m}^{k-1} \exp(w_0 + \sum_{i=1}^{n} w_{mi} x_i))$$

$$= \sum_{i=1}^{k} \left(\sum_{i=1}^{n} x_i\right) - \frac{\exp(w_0 + \sum_{i=1}^{n} w_{mi} x_i)}{1+\sum_{m}^{k-1} \exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)} \cdot \frac{d}{dw_i} (w_0 + \sum_{i=1}^{n} w_{mi} x_i).$$

$$= \sum_{i=1}^{k} \left(\sum_{i=1}^{n} x_i\right) - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)\right)}{1+\sum_{m}^{k-1} \exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)}.$$

$$= \sum_{i=1}^{k} \sum_{i=1}^{n} x_i \cdot \left(1 - \frac{\exp(w_{m0} + \sum_{i=1}^{n} w_{mi} \cdot x_i)}{1+\sum_{m}^{k-1} \exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)}\right).$$

2.d.

$$F(w_1 \ldots w_n) = \sum_{i=1}^{S} \ln(P(Y=y_i \mid X=x_i)) - \frac{\lambda}{2} \sum_{m=1}^{k-1} \sum_{i=1}^{n} w_{mi}^2$$

gradien F $= \frac{dF}{dw_i} = \frac{d}{dw_i} \sum_{i=1}^{k} \ln P(Y=y_i \mid X=x_i) - \frac{d}{dw_i} \frac{\lambda}{2} \sum_{m=1}^{k-1} \sum_{i=1}^{n} w_{mi}^2$

$$\frac{d}{dw_i} \frac{\lambda}{2} \sum_{m=1}^{k-1} \sum_{i=1}^{n} w_{mi}^2 = \frac{\lambda}{2} \sum_{m=1}^{k-1} \sum_{i=1}^{n} 2 w_{mi}$$

$$\frac{d}{dw_i} \sum_{i=1}^{S} \ln P(Y=y_i \mid X=x_i) = \sum_{i=1}^{S} \frac{d}{dw_i} \ln \frac{\exp(w_{k0} + \sum_{i=1}^{n} w_{ki} x_i)}{1 + \sum_{m=l}^{k-1} \exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)};$$

dari butir c didapat hasil turunannya adalah.

$$\frac{d}{dw_i} L(w_1 \ldots w_n) = \sum_{i=1}^{S} \sum_{i=1}^{n} x_i \cdot \left(1 - \frac{\exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)}{1 + \sum_{m=l}^{k-1} \exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)}\right).$$

sehingga didapat nilai dari $\frac{d}{dw_i} \sum_{i=1}^{S} \ln P(Y=y_i \mid X=x_i) - \frac{\lambda}{2} \sum_{m=1}^{k-1} \sum_{i=1}^{n} w_{mi}^2$

adalah :

$$\frac{dF}{dw_i} = \sum_{i=1}^{S} \sum_{i=1}^{n} x_i \cdot \left(1 - \frac{\exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)}{1 + \sum_{m=l}^{k-1} \exp(w_{m0} + \sum_{i=1}^{n} w_{mi} x_i)}\right) - \frac{\lambda}{2} \sum_{m=1}^{k-1} \sum_{i=1}^{n} 2 w_{mi}.$$