$$L(\omega) = -\frac{1}{2m}\left[\sum_{i=1}^{m} t^i \log\left(y(x^i,\omega)\right) + (1+t^i)\log\left(1-y(x^i,\omega)\right)\right]$$

$$+ \phi \sum_{j=1}^{N} \omega_j^2$$

where

ω : weight vector $(N+1, 1)$
x : input vector $(m, N+1)$
t^i : target
m : number of examples
N : number of features
$y(x^i,\omega)$: predicted value

From the provided python code (predict_y_prob()):

$y(x^i,\omega)$ is a sigmoid function (let $y(x^i,\omega) = \hat{y}^i$)

$$\hat{y} = \frac{1}{1+e^{-z}} \quad \& \quad z = \sum_{i=0}^{N} \omega_i \cdot x_i$$

$$\frac{\partial L(\omega)}{\partial \omega_k} = -\frac{1}{2m}\left[\sum_{i=1}^{m} \frac{\partial}{\partial \omega_k}\left(t^i \log(\hat{y}^i) + (1-t^i)\log(1-\hat{y}^i)\right)\right]$$

$$+ \frac{\partial}{\partial \omega_k} \phi \sum_{j=1}^{N} \omega_j^2$$

Firstly: $\frac{\partial}{\partial \omega_k} \phi \sum_{j=1}^{N} \omega_j^2$

we are differentiating a specific index of k

so $\phi(\omega_0^2 + \omega_1^2 + \dots \omega_k^2 \dots \omega_N^2)$

for the differentiation all other indexes are constants
Hence this would result in:

$$= 2\phi\omega_k$$

Second: $\frac{\partial}{\partial \omega_k}\Big[\underbrace{t^i \log(\hat{y}^i) + (1-t^i)\log(1-\hat{y}^i)}_{Q}\Big]$

But $\underset{(i)}{\frac{\partial Q}{\partial \omega_j}} = \underset{(ii)}{\frac{\partial Q}{\partial \hat{y}}} \cdot \underset{(iii)}{\frac{\partial \hat{y}^i}{\partial z}} \cdot \underset{(iv)}{\frac{\partial z}{\partial \omega}}$

$\therefore$ (ii) $\frac{\partial Q}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}}\left(t^i \log(\hat{y}^i) + (1-t^i)\log(1-\hat{y}^i)\right)$

$$= \frac{t^i}{\hat{y}^i} - \frac{1-t^i}{1-\hat{y}^i}$$

$$= \frac{t^i - t^i\hat{y}^i - \hat{y}^i + t^i\hat{y}^i}{\hat{y}(1-\hat{y}^i)} = \frac{t^i - \hat{y}^i}{\hat{y}^i(1-\hat{y}^i)}$$

(iii) $$\frac{\partial \hat{y}}{\partial z} = \frac{\partial \left(\frac{1}{1+e^{z}} \right)}{\partial z}$$

$$= \frac{-\left(-e^{-z}\right)}{(1+e^{-z})^2}$$

$$= +\left(\frac{1+e^{-z}-1}{(1+e^{-z})^2} \right)$$

$$= +\left(\frac{1+e^{-z}}{(1+e^{-z})^2} - \left(\frac{1}{1+e^{-z}} \right)^2 \right)$$

$$= +\hat{y}(1-\hat{y})$$

(iv) $$\frac{\partial z}{\partial w_k} = \frac{\partial}{\partial w_k} \sum_{j=0}^{N} w_j \cdot x_j^i$$

Similare to before:

$$(w_0 x_0^i + w_1 x_1^i + \dots + w_k x_k^i + \dots w_N x_N^i)$$

all w_j where $j \neq k$ can be considered as constants

$$\therefore \frac{\partial z}{\partial w_k} = \frac{\partial}{\partial w_k} (w_k x_k^i + \text{constant})$$

$$= x_k^i$$

$$\therefore \frac{\partial Q}{\partial \omega_k} = \frac{t^i - \hat{y}^i}{\hat{y}^i(1-\hat{y}^i)} \cdot \hat{y}^i(1-\hat{y}^i) \cdot x^i_k$$

$$\therefore \frac{\partial L}{\partial \omega_k} = -\frac{1}{2m}\left[\sum_{i=1}^{m} (t^i - \hat{y}^i)\cdot x^i_k\right] + 2\phi\,\omega_k$$

For the implementation :

at $k = 0$:

- Bias term not regularised
- $x^i_k = 1$

$$\therefore \omega_0 \mathrel{+}= \frac{\alpha}{2m}\sum_{i=1}^{m}(t^i - \hat{y}^i)$$

---

at $k \neq 0$:

$$\omega_k \mathrel{+}= \frac{\alpha}{2m}\left[\sum_{i=1}^{m}(t^i - \hat{y}^i)\cdot x^i_k\right] + 2\phi\,\omega_k\,\alpha$$