$$E(w) = \sum_{i=1}^{N} \log(1+e^{-w^T x_i y_i}) + \frac{\lambda}{2} ||w||^2$$

Gradient descent:

$$w_{j+1} = w_j - \alpha \frac{\delta E(w)}{\delta w_j} \qquad 0 \le j < m$$

$\frac{\partial E(w)}{\partial w}$ can be written in matrix form as

$$\frac{\delta E(w)}{\delta w} = \sum_{i=1}^{n} (\sigma(w^T \cdot x_i) - y_i) x_{ij} + \frac{\lambda \times \cancel{2} ||w||}{\cancel{2}}$$

$$= \frac{\delta E(w)}{\delta w} = \sum_{i=1}^{N} (\sigma(w^T \cdot x_i) - y_i) x_{ij} + \lambda ||w||$$

$$\Rightarrow \boxed{w_{j+1} = w_j - \alpha \sum_{i=1}^{N} (\sigma(w^T \cdot x_i) - y_i) x_{ij} + \lambda ||w||}$$

To be used for programming

for $j = 0$

$$w_0 = \alpha \sum_{i=1}^{N} (\sigma(w^T \cdot x_i) - y_i) x_{ij}$$

for $j = 1$ to n

$$w_{j+1} = w_j - \alpha \sum_{i=1}^{N} (\sigma(w^T \cdot x_i) - y_i) x_{ij} + \lambda ||w||$$