Solution 1.1

As we already know that update formulae for minimizing the loss using stochastic gradient descent with step size η is as follows :

for every iteration $\boldsymbol{w_{t+1}} = \boldsymbol{w_t} - \eta\, \boldsymbol{\nabla \mathcal{L}(w)}$

As given in question the loss function ,

$\mathcal{L}(w) = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \lambda||w||^2$ where $l(y_i, \hat{y}_i) = \log_e(1 + \exp(-y_i . \hat{y}_i))$ and $\hat{y}_i = \tanh(w, x_i)$

Now we need to find $\boldsymbol{\nabla \mathcal{L}(w)}$ **,** to find $\boldsymbol{\nabla \mathcal{L}(w)}$ we will partially differentiate $\mathcal{L}(w)$ w.r.t w.

Hence partial differentiation of $\mathcal{L}(w)$ w.r.t w is given as follow:

$$\nabla \mathcal{L}(w) = \frac{\partial(\sum_{i=1}^{N} l(-y_i, \hat{y}_i) + \lambda||w||^2)}{\partial w}$$

Now as given in question the on putting value of $l(y_i, \hat{y}_i)$ in above equation we get,

$$\nabla \mathcal{L}(w) = \frac{\partial\left(\sum_{i=1}^{N} \log_e(1 + \exp(-y_i . \tanh(w, x_i)))\right)}{\partial w} + 2\lambda w$$

$$\nabla \mathcal{L}(w) = \sum_{i=1}^{N} \frac{-\mathrm{y_i} . \mathrm{x_i} . \exp(-y_i . \tanh(w, x_i) . (1 - \tanh^2(w . x_i))}{1 + \exp(-y_i . \tanh(w, x_i)} + 2\lambda w$$

Hence the equation for w_{t+1} becomes

$$w_{t+1} = w_t - \eta \left(\frac{-\mathrm{y_i} . \mathrm{x_i} . \exp\,(-y_i . \tanh(w, x_i) . (1 - \tanh^2(w . x_i))}{1 + \exp\,(-y_i . \tanh(w, x_i))} + 2\lambda w \right)$$

$$\boxed{\boldsymbol{w_{t+1} = w_t(1 - 2\eta\lambda) + \eta \left(\frac{\mathrm{y_i} . \mathrm{x_i} . \exp\,(-y_i . \tanh(w_t, x_i) . (1 - \tanh^2(w_t . x_i))}{1 + \exp\,(-y_i . \tanh(w_t, x_i))} \right)}}$$