Model: $f : \mathbf{x} \rightarrow y,\ f(\mathbf{x}) = \begin{cases} 1, & \sigma(\mathbf{w}^T\mathbf{x}) > 0.5 \\ 0, & \text{otherwise} \end{cases}$, $\sigma(\eta) = \frac{1}{1+e^{-\eta}}$

Training data: $\mathcal{D} = \{(\mathbf{x_1}, y_1), \ldots, (\mathbf{x_N}, y_N)\}$

Evaluation through cross-entropy error (no regularization):
$\mathcal{E}(\mathbf{w}) = -\sum_{n=1}^{N} \left\{ y_n \log\left[\sigma(\mathbf{w}^T\mathbf{x_n})\right] + (1-y_n)\log\left[1-\sigma(\mathbf{w}^T\mathbf{x_n})\right] \right\}$

Evaluation through cross-entropy error ($l2$-norm regularization):
$\mathcal{E}(\mathbf{w}) = -\sum_{n=1}^{N} \left\{ y_n \log\left[\sigma(\mathbf{w}^T\mathbf{x_n})\right] + (1-y_n)\log\left[1-\sigma(\mathbf{w}^T\mathbf{x_n})\right] \right\} + \lambda \mathbf{w}^T\mathbf{w}$

$\nabla\mathcal{E}(\mathbf{w}) = \sum_{n=1}^{N} \left(\sigma(\mathbf{w}^T\mathbf{x_n}) - y_n\right)\mathbf{x_n} + 2\lambda\mathbf{w}$

Approximate solution through gradient descent:
$\mathbf{w}(k+1) := \mathbf{w} - \alpha\left(\sum_{n=1}^{N} \left(\sigma(\mathbf{w}^T\mathbf{x_n}) - y_n\right)\mathbf{x_n} + 2\lambda\mathbf{w}\right)$

$$\mathbf{H} = \nabla\left((\nabla\mathcal{E}(\mathbf{w}))^T\right) = \nabla\left(\sum_{n=1}^{N}\left(\sigma(\mathbf{w}^T\mathbf{x_n}) - y_n\right)\mathbf{x_n}^T + 2\lambda\mathbf{w}^T\right)$$
$$= \sum_{n=1}^{N} \underbrace{\sigma(\mathbf{w}^T\mathbf{x_n})}_{\in[0,1]} \cdot \underbrace{\left(1-\sigma(\mathbf{w}^T\mathbf{x_n})\right)}_{\in[0,1]} \cdot \underbrace{\left(\mathbf{x_n}\cdot\mathbf{x_n}^T\right)}_{\in\mathcal{R}^{D\times D}} + \lambda\mathbf{I}_{D\times D}$$

The above Hessian is positive semi-definite, since both matrices $\mathbf{x_n}\mathbf{x_n}^T$ and $\mathbf{I}_{D\times D}$ are positive semi-definite.