4.19 The Likelihood function for the two-class classification case will be similar to 4.89:

$$
\begin{aligned}
p(\mathbf{t}|\mathbf{w}) &= \prod_{n=1}^{N} y_n^{t_n}\{1-y_n\}^{1-t_n} \\
&= \prod_{n=1}^{N} \boldsymbol{\Phi}(a_n)^{t_n}\{1-\boldsymbol{\Phi}(a_n)\}^{1-t_n} \\
&= \prod_{n=1}^{N} \boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)^{t_n}\{1-\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)\}^{1-t_n}
\end{aligned}
$$

The log-likelihood is:

$$
\ln p(\mathbf{t}|\mathbf{w}) = \sum_{n=1}^{N} t_n \ln\left\{\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)\right\} + (1-t_n)\ln\left\{\{1-\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)\}\right\}
$$

The gradient of the log-likelihood is:

$$
\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w}) = \sum_{n=1}^{N} t_n \frac{1}{\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}\left(\frac{\partial\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\partial\mathbf{w}}\right) + (1-t_n)\frac{1}{1-\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}\left(-\frac{\partial\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\partial\mathbf{w}}\right)
$$

To find the gradient of the log-likelihood, we need to find the derivative of the activation function $\boldsymbol{\Phi}$.

$$
\begin{aligned}
\frac{\partial\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\partial\mathbf{w}} &= \frac{\partial}{\partial\mathbf{w}}\left(\frac{1}{2}\left\{1+\frac{1}{\sqrt{2}}erf(\mathbf{w}^T\boldsymbol{\phi}_n)\right\}\right) \\
&= \frac{\partial}{\partial\mathbf{w}}\left(\frac{1}{2}\left\{1+\frac{1}{\sqrt{2}}erf(\mathbf{w}^T\boldsymbol{\phi}_n)\right\}\right) \\
&= \frac{1}{2\sqrt{2}}\frac{\partial\, erf(\mathbf{w}^T\boldsymbol{\phi}_n)}{\partial\mathbf{w}} \\
&= \frac{1}{2\sqrt{2}}\frac{\partial}{\partial\mathbf{w}}\left(\frac{2}{\sqrt{\pi}}\int_0^{\mathbf{w}^T\boldsymbol{\phi}_n} \exp(-\theta^2/2)d\theta\right) \\
&= \frac{1}{\sqrt{2\pi}}\frac{\partial}{\partial\mathbf{w}}\left(\int_0^{\mathbf{w}^T\boldsymbol{\phi}_n} \exp(-\theta^2/2)d\theta\right)
\end{aligned}
$$

$$= \frac{1}{\sqrt{2\pi}} \left(\frac{\partial}{\partial \mathbf{w}^T \boldsymbol{\phi}_n} \left(\int_0^{\mathbf{w}^T \boldsymbol{\phi}_n} \exp(-\theta^2/2) d\theta \right) \right) \left(\frac{\partial \mathbf{w}^T \boldsymbol{\phi}_n}{\partial \mathbf{w}} \right)$$

Using this result, the derivative becomes:

$$= \frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T \boldsymbol{\phi}_n)^2/2) \boldsymbol{\phi}_n$$

Therefore, the gradient of the log-likelihood becomes:

$$\begin{aligned} \nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w}) = \sum_{n=1}^{N} \Bigg(& t_n \frac{1}{\boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)} \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T \boldsymbol{\phi}_n)^2/2) \boldsymbol{\phi}_n \right) \\ & + (1 - t_n) \frac{1}{(1 - \boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))} \left(-\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T \boldsymbol{\phi}_n)^2/2) \boldsymbol{\phi}_n \right) \Bigg) \end{aligned}$$

$$\begin{aligned} = \sum_{n=1}^{N} \Bigg(& \frac{t_n}{\boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)} \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T \boldsymbol{\phi}_n)^2/2) \boldsymbol{\phi}_n \right) \\ & + \frac{(t_n - 1)}{(1 - \boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))} \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T \boldsymbol{\phi}_n)^2/2) \boldsymbol{\phi}_n \right) \Bigg) \end{aligned}$$

$$= \sum_{n=1}^{N} \left(\frac{t_n}{\boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)} + \frac{(t_n - 1)}{(1 - \boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))} \right) \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T \boldsymbol{\phi}_n)^2/2) \boldsymbol{\phi}_n \right)$$

$$= \sum_{n=1}^{N} \left(\frac{t_n(1 - \boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)) + (t_n - 1)\boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)}{\boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)(1 - \boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))} \right) \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T \boldsymbol{\phi}_n)^2/2) \boldsymbol{\phi}_n \right)$$

$$= \sum_{n=1}^{N} \left(\frac{t_n - t_n \boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n) + t_n \boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n) - \boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)}{\boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)(1 - \boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))} \right) \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T \boldsymbol{\phi}_n)^2/2) \boldsymbol{\phi}_n \right)$$

$$\Longrightarrow \nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w}) = \sum_{n=1}^{N} \left(\frac{t_n - \boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)}{\boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)(1 - \boldsymbol{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))} \right) \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T \boldsymbol{\phi}_n)^2/2) \boldsymbol{\phi}_n \right)$$

Finding the Hessian of the log-likelihood:

$$\mathbf{H} = \nabla\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w}) = \sum_{n=1}^{N} \left(\left(\frac{t_n - \boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))} \right) \frac{\partial}{\partial \mathbf{w}} \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2)\boldsymbol{\phi}_n \right) \right.$$
$$\left. + \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2)\boldsymbol{\phi}_n \right) \frac{\partial}{\partial \mathbf{w}} \left(\frac{t_n - \boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))} \right) \right)$$

Solving the derivative in the first term:

$$\frac{\partial}{\partial \mathbf{w}} \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2)\boldsymbol{\phi}_n \right)$$
$$= \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2) \right) \frac{\partial(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2\boldsymbol{\phi}_n/2)}{\partial \mathbf{w}}$$
$$= \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2) \right) \left(\frac{-1}{2} \right) (2\mathbf{w}^T\boldsymbol{\phi}_n) \frac{\partial(\mathbf{w}^T\boldsymbol{\phi}_n\boldsymbol{\phi}_n)}{\partial \mathbf{w}}$$
$$= \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2) \right) \left(\frac{-1}{2} \right) (2\mathbf{w}^T\boldsymbol{\phi}_n) \frac{\partial(\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T\mathbf{w})}{\partial \mathbf{w}}$$
$$= -\left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2) \right) (\mathbf{w}^T\boldsymbol{\phi}_n)\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T$$

Solving the derivative in the second term:

$$\frac{\partial}{\partial \mathbf{w}} \left(\frac{t_n - \boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))} \right)$$
$$= \frac{\partial}{\partial \boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)} \left(\frac{t_n - \boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))} \right) \frac{\partial \boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\partial \mathbf{w}}$$

$$\frac{\partial}{\partial \boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)} \left(\frac{t_n - \boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))} \right)$$
$$= \frac{\partial}{\partial \boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)} \left(\frac{t_n}{\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))} - \frac{1}{(1-\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))} \right)$$

$$= \frac{\partial}{\partial \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)} \left(t_n (\mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)(1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)))^{-1} - (1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))^{-1} \right)$$

$$= t_n \frac{\partial (\mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)(1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)))^{-1}}{\partial \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)} - \frac{\partial (1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))^{-1}}{\partial \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)}$$

$$= t_n (-1)(\mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)(1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)))^{-2} \left(\frac{\partial \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)(1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))}{\partial \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)} \right)$$
$$-(-1)(1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))^{-2} \left(\frac{\partial (1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))}{\partial \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)} \right)$$

$$= t_n (-1)(\mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)(1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)))^{-2} \left((1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)) + \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)(-1) \right)$$
$$-(-1)(1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))^{-2}(-1)$$

$$= -t_n (\mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)(1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)))^{-2} \left(1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n) - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n) \right)$$
$$-(1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))^{-2}$$

$$= -\frac{1}{(1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))^2} - \frac{t_n \left(1 - 2\mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n) \right)}{(\mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)(1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)))^2}$$

$$= -\left(\frac{(\mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n))^2 + t_n \left(1 - 2\mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n) \right)}{(\mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)(1 - \mathbf{\Phi}(\mathbf{w}^T \boldsymbol{\phi}_n)))^2} \right)$$

Substituting into the expression for the Hessian,

$$\mathbf{H} = \nabla\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w})$$
$$= \sum_{n=1}^{N} \left(\left(\frac{t_n - \mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))} \right) \left(-\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2) \right) (\mathbf{w}^T\boldsymbol{\phi}_n)\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T \right.$$
$$\left. - \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2)\boldsymbol{\phi}_n \right) \left(\frac{t_n \left(1 - 2\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)\right) + (\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))^2}{(\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)))^2} \right) \frac{\partial \mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\partial \mathbf{w}} \right)$$

$$\nabla\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w}) = \sum_{n=1}^{N} \left(\left(\frac{t_n - \mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))} \right) \right.$$
$$\left(-\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2) \right) (\mathbf{w}^T\boldsymbol{\phi}_n)\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T$$
$$- \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2) \right)$$
$$\left(\frac{t_n \left(1 - 2\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)\right) + (\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))^2}{(\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)))^2} \right)$$
$$\left. \frac{\partial(\boldsymbol{\phi}_n\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))}{\partial \mathbf{w}} \right)$$

$$\nabla\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w}) = \sum_{n=1}^{N} \left(\left(\frac{t_n - \mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))} \right) \right.$$
$$\left(-\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2) \right) (\mathbf{w}^T\boldsymbol{\phi}_n)\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T$$
$$- \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2) \right)$$
$$\left(\frac{t_n \left(1 - 2\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)\right) + (\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))^2}{(\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\mathbf{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)\boldsymbol{\phi}_n^T))^2} \right)$$
$$\left. \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2)\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T \right) \right)$$

$$
\begin{aligned}
&\nabla\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w}) \\
&= -\sum_{n=1}^{N} \left(\frac{t_n - \boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)}{\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))} (\mathbf{w}^T\boldsymbol{\phi}_n) \right. \\
&\qquad + \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2) \right) \\
&\qquad \left. \left(\frac{t_n \left(1 - 2\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)\right) + (\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n))^2}{(\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)(1-\boldsymbol{\Phi}(\mathbf{w}^T\boldsymbol{\phi}_n)))^2} \right) \right) \\
&\qquad \left(\frac{1}{\sqrt{2\pi}} \exp(-(\mathbf{w}^T\boldsymbol{\phi}_n)^2/2) \right) \boldsymbol{\phi}_n \boldsymbol{\phi}_n^T
\end{aligned}
$$