

4.13 The error function is given by 4.90:

$$E(\mathbf{w}) = - \sum_{n=1}^N \{t_n \ln y_n + (1 - t_n) \ln(1 - y_n)\}$$

From 4.87, we know that $y = \sigma(\mathbf{w}^T \phi)$. Substituting for y , we get:

$$E(\mathbf{w}) = - \sum_{n=1}^N \{t_n \ln(\sigma(\mathbf{w}^T \phi_n)) + (1 - t_n) \ln(1 - \sigma(\mathbf{w}^T \phi_n))\}$$

Taking the gradient of the error function with respect to \mathbf{w} , we obtain:

$$\begin{aligned} \frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} &= - \frac{\partial}{\partial \mathbf{w}} \sum_{n=1}^N \{t_n \ln(\sigma(\mathbf{w}^T \phi_n)) + (1 - t_n) \ln(1 - \sigma(\mathbf{w}^T \phi_n))\} \\ &= - \sum_{n=1}^N \left\{ \frac{\partial t_n \ln(\sigma(\mathbf{w}^T \phi_n))}{\partial \mathbf{w}} + \frac{\partial (1 - t_n) \ln(1 - \sigma(\mathbf{w}^T \phi_n))}{\partial \mathbf{w}} \right\} \end{aligned}$$

Evaluating the derivative of the first term inside the summation:

$$\begin{aligned} &\frac{\partial t_n \ln(\sigma(\mathbf{w}^T \phi_n))}{\partial \mathbf{w}} \\ &= t_n \frac{\partial \ln(\sigma(\mathbf{w}^T \phi_n))}{\partial \mathbf{w}} \\ &= t_n \left(\frac{\partial \ln(\sigma(\mathbf{w}^T \phi_n))}{\partial \sigma(\mathbf{w}^T \phi_n)} \right) \left(\frac{\partial \sigma(\mathbf{w}^T \phi_n)}{\partial (\mathbf{w}^T \phi_n)} \right) \left(\frac{\partial (\mathbf{w}^T \phi_n)}{\partial \mathbf{w}} \right) \\ &= t_n \left(\frac{1}{\sigma(\mathbf{w}^T \phi_n)} \right) (\sigma(\mathbf{w}^T \phi_n)(1 - \sigma(\mathbf{w}^T \phi_n))) (\phi_n) \\ &= t_n (1 - \sigma(\mathbf{w}^T \phi_n)) (\phi_n) \\ &= t_n \phi_n - t_n \sigma(\mathbf{w}^T \phi_n) \phi_n \end{aligned}$$

Evaluating the derivative of the second term inside the summation:

$$\begin{aligned}
& \frac{\partial(1 - t_n) \ln(1 - \sigma(\mathbf{w}^T \phi_n))}{\partial \mathbf{w}} \\
&= (1 - t_n) \frac{\partial \ln(1 - \sigma(\mathbf{w}^T \phi_n))}{\partial \mathbf{w}} \\
&= (1 - t_n) \left(\frac{\partial \ln(1 - \sigma(\mathbf{w}^T \phi_n))}{\partial(1 - \sigma(\mathbf{w}^T \phi_n))} \right) \left(\frac{\partial(1 - \sigma(\mathbf{w}^T \phi_n))}{\partial(\mathbf{w}^T \phi_n)} \right) \left(\frac{\partial(\mathbf{w}^T \phi_n)}{\partial \mathbf{w}} \right) \\
&= (1 - t_n) \left(\frac{1}{(1 - \sigma(\mathbf{w}^T \phi_n))} \right) (-\sigma(\mathbf{w}^T \phi_n)(1 - \sigma(\mathbf{w}^T \phi_n))) (\phi_n) \\
&= (1 - t_n) (-\sigma(\mathbf{w}^T \phi_n)) (\phi_n) \\
&= t_n \sigma(\mathbf{w}^T \phi_n) \phi_n - \sigma(\mathbf{w}^T \phi_n) \phi_n
\end{aligned}$$

Substituting these derivatives back into the error function, we get:

$$\begin{aligned}
E(\mathbf{w}) &= - \sum_{n=1}^N \{t_n \phi_n - t_n \sigma(\mathbf{w}^T \phi_n) \phi_n + t_n \sigma(\mathbf{w}^T \phi_n) \phi_n - \sigma(\mathbf{w}^T \phi_n) \phi_n\} \\
&= - \sum_{n=1}^N \{t_n \phi_n - \sigma(\mathbf{w}^T \phi_n) \phi_n\} \\
&= - \sum_{n=1}^N \{t_n - \sigma(\mathbf{w}^T \phi_n)\} \phi_n \\
&= - \sum_{n=1}^N \{t_n - y_n\} \phi_n \\
&= \sum_{n=1}^N \{y_n - t_n\} \phi_n
\end{aligned}$$

which is the same as 4.91.