4.14 If a dataset is linearly separable, then the decision boundary can perfectly separate the classes. This means that for $y_n = 1$, $(\mathbf{w}^T\boldsymbol{\phi}_n) > 0$ and for $y_n = 0$, $(\mathbf{w}^T\boldsymbol{\phi}_n) < 0$.

For $t_n = 1$, the error $= -\ln y_n = -\ln\left(\frac{1}{1+\exp(-\mathbf{w}^T\boldsymbol{\phi}_n)}\right) = \ln\big(1+\exp(-\mathbf{w}^T\boldsymbol{\phi}_n)\big)$.

This error can be minimized if $\exp(-\mathbf{w}^T\boldsymbol{\phi}_n) \to -\infty$, $\implies \mathbf{w}^T\boldsymbol{\phi}_n \to \infty$.

For $t_n = 0$, the error $= -(1 - \ln y_n) = -\left(1 - \ln\left(\frac{1}{1+\exp(-\mathbf{w}^T\boldsymbol{\phi}_n)}\right)\right) = -1 - \ln\big(1+\exp(-\mathbf{w}^T\boldsymbol{\phi}_n)\big)$.

This error can be minimized if $\exp(-\mathbf{w}^T\boldsymbol{\phi}_n) \to \infty$, $\implies \mathbf{w}^T\boldsymbol{\phi}_n \to -\infty$.