6.16 The gradient of f is:

$$\nabla_{\mathbf{w}} J = \begin{bmatrix} \frac{\partial f}{\partial w_1} + \frac{g(\mathbf{w}^T\mathbf{w})}{\partial w_1} \\ \frac{\partial f}{\partial w_2} + \frac{g(\mathbf{w}^T\mathbf{w})}{\partial w_2} \\ \vdots \\ \frac{\partial f}{\partial w_M} + \frac{g(\mathbf{w}^T\mathbf{w})}{\partial w_M} \end{bmatrix}$$

$$\begin{aligned} \frac{\partial f}{\partial w_m} &= \sum_{n=1}^{N} \frac{\partial f}{\partial \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_n)} \cdot \frac{\partial \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_n)}{\partial w_m} \\ &= \sum_{n=1}^{N} \frac{\partial f}{\partial \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_n)} \cdot \phi_m(\mathbf{x}_n) \\ &= (\nabla_{(\boldsymbol{\Phi}\mathbf{w})} f)^T \boldsymbol{\phi}_m \end{aligned}$$

$$\begin{aligned} \frac{g(\mathbf{w}^T\mathbf{w})}{\partial w_m} &= \frac{g(\mathbf{w}^T\mathbf{w})}{\partial \mathbf{w}^T\mathbf{w}} \cdot \frac{\partial \mathbf{w}^T\mathbf{w}}{\partial w_m} \\ &= g'(\mathbf{w}^T\mathbf{w}) \cdot 2w_m \end{aligned}$$

Setting the gradient to 0, we get:

$$(\nabla_{(\boldsymbol{\Phi}\mathbf{w})} f)^T \boldsymbol{\phi}_m + 2g'(\mathbf{w}^T\mathbf{w})w_m = 0$$

$$\implies w_m = -\frac{(\nabla_{(\boldsymbol{\Phi}\mathbf{w})} f)^T \boldsymbol{\phi}_m}{2g'(\mathbf{w}^T\mathbf{w})}$$

$$\implies \mathbf{w} = -\frac{1}{2g'(\mathbf{w}^T\mathbf{w})} \begin{bmatrix} ---\boldsymbol{\phi}_1 --- \\ ---\boldsymbol{\phi}_2 --- \\ \vdots \\ ---\boldsymbol{\phi}_M --- \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial \mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_1)} \\ \frac{\partial f}{\partial \mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_2)} \\ \vdots \\ \frac{\partial f}{\partial \mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_N)} \end{bmatrix}$$

$$= -\frac{1}{2g'(\mathbf{w}^T\mathbf{w})} \sum_{n=1}^{N} \boldsymbol{\phi}(\mathbf{x}_n) \cdot \frac{\partial f}{\partial \mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_n)}$$

$$\implies \alpha_n = -\frac{1}{2g'(\mathbf{w}^T\mathbf{w})} \cdot \frac{\partial f}{\partial \mathbf{w}^T\boldsymbol{\phi}(\mathbf{x}_n)}$$

and

$$\mathbf{w}_\perp = \mathbf{0}$$