Finding the Compact Derivative

Let's first review some key notations, remember how we defined a function?

$$f(x) = w_1\phi_1(x) + w_2\phi_2(x) + w_3\phi_3(x) + ... = \underbrace{\begin{bmatrix} w_1 & w_2 & w_3 & ... \end{bmatrix}}_{w^\top} \underbrace{\begin{bmatrix} \phi_1(x) \\ \phi_2(x) \\ ... \end{bmatrix}}_{\phi(x)} = \underbrace{w^\top \phi(x) = \phi(x)^\top w.}_{\text{they are equivalent}}$$

Next, if we have n samples consists of $\{x_1, x_2, x_3, ...\}$, then our prediction of each sample would be

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \\ ... \end{bmatrix} = \begin{bmatrix} \phi(x_1)^\top w \\ \phi(x_2)^\top w \\ \phi(x_3)^\top w \\ ... \end{bmatrix} = \underbrace{\begin{bmatrix} \phi(x_1)^\top \\ \phi(x_2)^\top \\ \phi(x_3)^\top \\ ... \end{bmatrix}}_{\text{Let's call it } \Phi} w = \Phi w.$$

The key takeaway is that everyone of our prediction can be denoted simply as Φw and that

$$\Phi = \begin{bmatrix} \phi(x_1)^\top \\ \phi(x_2)^\top \\ \phi(x_3)^\top \\ ... \end{bmatrix}$$

Given that we are solving a regression problem, the derivative can be written as

$$\mathcal{L}(w) = \frac{1}{2}\sum_{i}^{N}(w^\top \phi(x_i) - y_i)^2 \tag{4}$$

$$\frac{d\mathcal{L}}{dw} = \sum_{i}^{N}(w^\top \phi(x_i) - y_i)\phi(x_i) \tag{5}$$

$$= (w^\top \phi(x_1) - y_1)\phi(x_1) + (w^\top \phi(x_2) - y_2)\phi(x_2) + (w^\top \phi(x_3) - y_3)\phi(x_3) + \dots \tag{6}$$

You can find $\frac{d\mathcal{L}}{dw}$ using this equation. But this is way too slow. Let's see a faster trick. Let

$$a_i = (w^\top \phi(x_i) - y_i) = (\phi(x_i)^\top w - y_i) \tag{7}$$

then

$$\frac{d\mathcal{L}}{dw} = a_1\phi(x_1) + a_2\phi(x_2) + a_3\phi(x_3) + a_4\phi(x_4) \tag{8}$$

$$= \begin{bmatrix}\phi(x_1) & \phi(x_2) & \phi(x_3) & \phi(x_4)\end{bmatrix}\begin{bmatrix}a_1\\ a_2\\ a_3\\ a_4\end{bmatrix} \tag{9}$$

$$= \begin{bmatrix}\phi(x_1) & \phi(x_2) & \phi(x_3) & \phi(x_4)\end{bmatrix}\begin{bmatrix}(\phi(x_1)^\top w - y_1)\\ (\phi(x_2)^\top w - y_2)\\ (\phi(x_3)^\top w - y_3)\\ (\phi(x_4)^\top w - y_4)\end{bmatrix} \tag{10}$$

$$= \begin{bmatrix}\phi(x_1) & \phi(x_2) & \phi(x_3) & \phi(x_4)\end{bmatrix}\left[\begin{bmatrix}\phi(x_1)^\top\\ \phi(x_2)^\top\\ \phi(x_3)^\top\\ \phi(x_4)^\top\end{bmatrix} w - \begin{bmatrix}y_1\\ y_2\\ y_3\\ y_4\end{bmatrix}\right] \tag{11}$$

$$= \Phi^\top \left[\Phi w - y\right] \tag{12}$$

Last page, take away

$$\Phi = \begin{bmatrix}\phi(x_1)^\top\\ \phi(x_2)^\top\\ \phi(x_3)^\top\\ \dots\end{bmatrix}$$

Much easier to calculate one vs the other