Kernelized SVM Hypothesis

First, note that when transforming the features of the inputs, our definition of $\mathbf{w}$ becomes

$$\mathbf{w} = \sum_{i=1}^{N} \alpha^{(i)} y^{(i)} \phi(\mathbf{x}^{(i)})$$

since we are still trying to find $\arg\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2$, which, when using Lagrange multipliers to include the constraints, becomes

$$\mathcal{L}(\mathbf{w}, w_0, \alpha) = \frac{1}{2}\mathbf{w}^\top \mathbf{w} + \sum_{i=1}^{N} \alpha^{(i)}(1 - y^{(i)}(\mathbf{w}^\top \phi(\mathbf{x}^{(i)}) + w_0))$$

and when deriving this with respect to $\mathbf{w}$ and setting the derivative to 0, we obtain

$$\frac{\partial}{\partial \mathbf{w}}\mathcal{L} = w - \sum_{i=1}^{N} \alpha^{(i)} y^{(i)} \phi(\mathbf{x}^{(i)}) = 0 \iff \mathbf{w} = \sum_{i=1}^{N} \alpha^{(i)} y^{(i)} \phi(\mathbf{x}^{(i)})$$

We can now manipulate the hypothesis h to obtain its "kernelized" form. Let's start by substituting for $\mathbf{w}$, thus

$$\begin{aligned} h(\mathbf{x}) &= \operatorname{sgn}\left(w_0 + \phi(\mathbf{x})^\top \mathbf{w}\right) \\ &= \operatorname{sgn}\left(w_0 + \phi(\mathbf{x})^\top \left(\sum_{i=1}^{N} \alpha^{(i)} y^{(i)} \phi(\mathbf{x}^{(i)})\right)\right) \end{aligned}$$

Since $\phi(\mathbf{x})^\top$ is a linear map, by the property of additivity, we have

$$h(\mathbf{x}) = \operatorname{sgn}\left(w_0 + \sum_{i=1}^{N} \phi(\mathbf{x})^\top \alpha^{(i)} y^{(i)} \phi(\mathbf{x}^{(i)})\right)$$

Since $\alpha^{(i)}$ and $y^{(i)}$ are scalars, by the property of homogeneity, we then have

$$h(\mathbf{x}) = \operatorname{sgn}\left(w_0 + \sum_{i=1}^{N} \alpha^{(i)} y^{(i)} \phi(\mathbf{x})^\top \phi(\mathbf{x}^{(i)})\right)$$

Since $\phi(\mathbf{x})^\top \phi(\mathbf{x}^{(i)}) = \langle \phi(\mathbf{x}), \phi(\mathbf{x}^{(i)}) \rangle$ and $\langle \phi(\mathbf{x}), \phi(\mathbf{x}^{(i)}) \rangle = \langle \phi(\mathbf{x}^{(i)}), \phi(\mathbf{x}) \rangle$ because $\phi(\mathbf{x}^{(j)}) \in \mathbb{R}^D$, we have

$$\begin{aligned} h(\mathbf{x}) &= \operatorname{sgn}\left(w_0 + \sum_{i=1}^{N} \alpha^{(i)} y^{(i)} \langle \phi(\mathbf{x}), \phi(\mathbf{x}^{(i)}) \rangle\right) \\ &= \operatorname{sgn}\left(w_0 + \sum_{i=1}^{N} \alpha^{(i)} y^{(i)} \langle \phi(\mathbf{x}^{(i)}), \phi(\mathbf{x}) \rangle\right) \end{aligned}$$

And by the definition of the kernel function, we finally have that

$$h(\mathbf{x}) = \operatorname{sgn}\left(w_0 + \sum_{i=1}^{N} \alpha^{(i)} y^{(i)} k(\mathbf{x}^{(i)}, \mathbf{x})\right)$$