Kernel Trick in Linear Regression

In standard linear regression, the output is:

$$y = \mathbf{X}\mathbf{w}$$

where $\mathbf{X}$ is the feature vector, and $\mathbf{w}$ is the weight vector.

For nonlinear regression, we map the data into a higher-dimensional space using a function $\Phi(\mathbf{X})$:

$$y = \Phi(\mathbf{X})\mathbf{w}$$

where $\Phi(\mathbf{X})$ is the nonlinear feature map.

The kernel trick allows us to compute the dot product in this higher-dimensional space without explicitly mapping the data. This is achieved by using a kernel function $K(\mathbf{X}, \mathbf{X}')$, where:

$$K(\mathbf{X}, \mathbf{X}') = \langle \Phi(\mathbf{X}), \Phi(\mathbf{X}') \rangle$$

So, the weights are computed by the kernel matrix K as:

$$\mathbf{w} = \mathbf{K}\mathbf{a}$$

where $\mathbf{K} = \Phi(\mathbf{X})\Phi(\mathbf{X})^T$ and $\mathbf{a}$ is the vector of dual coefficients.