Thus the map κ given by $\kappa(x, y) = (\kappa_1(x, y))^2$ is a kernel map associated with the feature map $\varphi\colon X \to \mathbb{R}^n \times \mathbb{R}^n$. The feature map φ is a direct generalization of the feature map φ_2 of Example 53.2.

The above argument is immediately adapted to show that if $\varphi_1\colon X \to \mathbb{R}^{n_1}$ and $\varphi_2\colon X \to \mathbb{R}^{n_2}$ are two feature maps and if $\kappa_1(x, y) = \langle \varphi_1(x), \varphi_1(y)\rangle$ and $\kappa_2(x, y) = \langle \varphi_2(x), \varphi_2(y)\rangle$ are the corresponding kernel functions, then the map defined by

$$\kappa(x, y) = \kappa_1(x, y)\kappa_2(x, y)$$

is a kernel function for the feature space $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ and the feature map

$$\varphi(x)_{(i,j)} = (\varphi_1(x))_i(\varphi_2(x))_j, \qquad 1 \leq i \leq n_1,\ 1 \leq j \leq n_2.$$

Example 53.4. Note that the feature map $\varphi\colon X \to \mathbb{R}^n \times \mathbb{R}^n$ is not very economical because if $i \neq j$ then the components $\varphi_{(i,j)}(x)$ and $\varphi_{(j,i)}(x)$ are both equal to $(\varphi_1(x))_i(\varphi_1(x))_j$. Therefore we can define the more economical embedding $\varphi'\colon X \to \mathbb{R}^{\binom{n+1}{2}}$ given by

$$\varphi'(x)_{(i,j)} = \begin{cases} (\varphi_1(x))_i^2 & i = j, \\ \sqrt{2}(\varphi_1(x))_i(\varphi_1(x))_j & i < j, \end{cases}$$

where the pairs (i, j) with $1 \leq i \leq j \leq n$ are ordered lexicographically. The feature map φ is a direct generalization of the feature map φ_1 of Example 53.2.

Observe that φ' can also be defined in the following way which makes it easier to come up with the generalization to any power:

$$\varphi'_{(i_1,\ldots,i_n)}(x) = \binom{2}{i_1 \cdots i_n}^{1/2} (\varphi_1(x))_1^{i_1}(\varphi_1(x))_1^{i_2} \cdots (\varphi_1(x))_1^{i_n}, \quad i_1 + i_2 + \cdots + i_n = 2,\ i_j \in \mathbb{N},$$

where the n-tuples $(i_1, \ldots, i_n)$ are ordered lexicographically. Recall that for any $m \geq 1$ and any $(i_1, \ldots, i_n) \in \mathbb{N}^m$ such that $i_1 + i_2 + \cdots + i_n = m$, we have

$$\binom{m}{i_1 \cdots i_n} = \frac{m!}{i_1! \cdots i_n!}.$$

More generally, for any $m \geq 2$, using the multinomial theorem, we can define a feature embedding $\varphi\colon X \to \mathbb{R}^{\binom{n+m-1}{m}}$ defining the kernel function κ given by $\kappa(x, y) = (\kappa_1(x, y))^m$, with φ given by

$$\varphi_{(i_1,\ldots,i_n)}(x) = \binom{m}{i_1 \cdots i_n}^{1/2} (\varphi_1(x))_1^{i_1}(\varphi_1(x))_1^{i_2} \cdots (\varphi_1(x))_1^{i_n}, \quad i_1 + i_2 + \cdots + i_n = m,\ i_j \in \mathbb{N},$$

where the n-tuples $(i_1, \ldots, i_n)$ are ordered lexicographically.