

# Pattern Recognition and Machine Learning

## Cristopher Bishop

### Exercise Solutions

Stefan Stefanache

April 7, 2022

# Chapter 1

## Kernel Methods

### Exercise 6.1 ★★

Consider the dual formulation of the least squares linear regression problem given in Section 6.1. Show that the solution for the components  $a_n$  of the vector  $\mathbf{a}$  can be expressed as a linear combination of the elements of the vector  $\phi(\mathbf{x}_n)$ . Denoting these coefficients by the vector  $\mathbf{w}$ , show that the dual of the dual formulation is given by the original representation in terms of the parameter vector  $\mathbf{w}$ .

*Proof.* By rewriting (6.4), one has that

$$\begin{aligned} a_n &= -\frac{1}{\lambda} \{ \mathbf{w}^T \phi(\mathbf{x}_n) - t_n \} \\ &= -\frac{1}{\lambda} \left\{ \sum_{i=1}^M w_i \phi_i(\mathbf{x}_n) - \frac{t_n}{\sum_{i=1}^M \phi_i(\mathbf{x}_n)} \sum_{i=1}^M \phi_i(\mathbf{x}_n) \right\} \\ &= \sum_{i=1}^M \left( \frac{t_n}{\lambda \sum_{i=1}^M \phi_i(\mathbf{x}_n)} - \frac{w_i}{\lambda} \right) \phi_i(\mathbf{x}_n) \\ &= \sum_{i=1}^M \Omega_{ni} \phi_i(\mathbf{x}_n) \\ &= \Omega_n^T \phi(\mathbf{x}_n) \end{aligned}$$

where

$$\Omega_{ni} = \frac{t_n}{\lambda \sum_{i=1}^M \phi_i(\mathbf{x}_n)} - \frac{w_i}{\lambda}$$

Therefore,  $a_n$  can be written as a linear combination of the elements of  $\phi(\mathbf{x}_n)$  and

$$\mathbf{a} = \text{diag}(\Omega \Phi)$$

□

### Exercise 6.3 ★

The nearest-neighbour classifier (Section 2.5.2) assigns a new input vector  $\mathbf{x}$  to the same class as that of the nearest input vector  $\mathbf{x}_n$  from the training set, where in the simple case, the distance

is defined by the Euclidean metric  $\|\mathbf{x} - \mathbf{x}_n\|^2$ . By expressing this rule in terms of scalar product and then making use of kernel substitution, formulate the nearest-neighbour classifier for a general nonlinear kernel.

*Proof.* Since we're dealing with inner products over  $\mathbb{R}$ , the Euclidian metric can be rewritten as

$$\|\mathbf{x} - \mathbf{x}_n\|^2 = \langle \mathbf{x} - \mathbf{x}_n, \mathbf{x} - \mathbf{x}_n \rangle = \langle \mathbf{x}, \mathbf{x} \rangle - 2\langle \mathbf{x}, \mathbf{x}_n \rangle + \langle \mathbf{x}_n, \mathbf{x}_n \rangle$$

Similarly to what happens in Section 6.2, using kernel substitution above to replace  $\langle \mathbf{x}, \mathbf{x}' \rangle$  with a nonlinear kernel  $\kappa(\mathbf{x}, \mathbf{x}')$  yields the nearest-neighbour classifier for a general nonlinear kernel:

$$k(\mathbf{x}, \mathbf{x}') = \kappa(\mathbf{x}, \mathbf{x}) - 2\kappa(\mathbf{x}, \mathbf{x}_n) + \kappa(\mathbf{x}_n, \mathbf{x}_n)$$

□

## Exercise 6.4 ★

In Appendix C, we give an example of a matrix that has positive elements but that has a negative eigenvalue and hence that is not positive definite. Find an example of the converse property, namely a  $2 \times 2$  matrix with positive eigenvalues that has at least one negative element.

*Proof.* Consider the matrix

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$$

$A$  contains one negative element and the eigenvalues of  $A$  are  $\lambda_1 = 1$  and  $\lambda_2 = 3$ , which proves that a matrix can be positive definite and have negative elements. □

## Exercise 6.5 ★

Verify the results (6.13) and (6.14) for constructing valid kernels.

*Proof.* Since  $k_1$  is a valid kernel, let  $\boldsymbol{\alpha}$  be a feature mapping such that

$$k_1(\mathbf{x}, \mathbf{x}') = \langle \boldsymbol{\alpha}(\mathbf{x}), \boldsymbol{\alpha}(\mathbf{x}') \rangle$$

Using the fact that an inner product on a real vector space is a positive-definite symmetric bilinear form, we have that

$$ck_1(\mathbf{x}, \mathbf{x}') = c\langle \boldsymbol{\alpha}(\mathbf{x}), \boldsymbol{\alpha}(\mathbf{x}') \rangle = \langle \sqrt{c}\boldsymbol{\alpha}(\mathbf{x}), \sqrt{c}\boldsymbol{\alpha}(\mathbf{x}') \rangle = \langle \boldsymbol{\beta}(\mathbf{x}), \boldsymbol{\beta}(\mathbf{x}') \rangle$$

where  $c > 0$  is a constant and  $\boldsymbol{\beta}(\mathbf{x}) = \sqrt{c}\boldsymbol{\alpha}(\mathbf{x})$ . Therefore, the new kernel

$$k(\mathbf{x}, \mathbf{x}') = ck_1(\mathbf{x}, \mathbf{x}') \tag{6.13}$$

is valid. Analogously, since  $f(\cdot)$  is a real-valued function,

$$f(\mathbf{x})k_1(\mathbf{x}, \mathbf{x}')f(\mathbf{x}') = f(\mathbf{x})\langle \boldsymbol{\alpha}(\mathbf{x}), \boldsymbol{\alpha}(\mathbf{x}') \rangle f(\mathbf{x}') = \langle f(\mathbf{x})\boldsymbol{\alpha}(\mathbf{x}), f(\mathbf{x}')\boldsymbol{\alpha}(\mathbf{x}') \rangle = \langle \boldsymbol{\gamma}(\mathbf{x}), \boldsymbol{\gamma}(\mathbf{x}') \rangle$$

where  $\boldsymbol{\gamma}(\mathbf{x}) = f(\mathbf{x})\boldsymbol{\alpha}(\mathbf{x})$ . As a result, the kernel

$$k(\mathbf{x}, \mathbf{x}') = f(\mathbf{x})k_1(\mathbf{x}, \mathbf{x}')f(\mathbf{x}') \tag{6.14}$$

will also be valid. □

## Exercise 6.6 ★

Verify the results (6.15) and (6.16) for constructing valid kernels.

*Proof.* Let  $q(\cdot)$  be a polynomial with nonnegative coefficients. Since in the polynomial kernels are summed and multiplied by nonnegative constants or other kernels, combining (6.13), (6.17) and (6.18) proves that the kernel

$$k(\mathbf{x}, \mathbf{x}') = q(k_1(\mathbf{x}, \mathbf{x}')) \quad (6.15)$$

is valid. Now, the exponential function is defined as

$$\exp(x) := \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

, so

$$\exp(k_1(\mathbf{x}, \mathbf{x}')) = \sum_{i=0}^{\infty} \frac{k_1(\mathbf{x}, \mathbf{x}')^i}{i!}$$

Note that the exponential of a kernel is an infinite sequence of kernel sums and products (with itself or nonnegative constants), so by using (6.13), (6.17), (6.18) again, one has that the new kernel

$$k(\mathbf{x}, \mathbf{x}') = \exp(k_1(\mathbf{x}, \mathbf{x}')) \quad (6.16)$$

is valid.  $\square$

## Exercise 6.7 ★

Verify the results (6.17) and (6.18) for constructing valid kernels.

*Proof.* Let  $K_1$  and  $K_2$  be the Gram matrices corresponding to the kernels  $k_1$  and  $k_2$ . Therefore, they are positive semidefinite matrices, so for any  $\mathbf{a} \in \mathbb{R}^n$ , one has that

$$\mathbf{a}^T \mathbf{H} \mathbf{a} = \mathbf{a}^T (\mathbf{H}_1 + \mathbf{H}_2) \mathbf{a} = \mathbf{a}^T \mathbf{H}_1 \mathbf{a} + \mathbf{a}^T \mathbf{H}_2 \mathbf{a} > 0$$

Since  $\mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2$  is positive semidefinite and corresponds to the Gram matrix of the kernel  $k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}')$ , one has that the kernel

$$k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}') \quad (6.17)$$

is valid. Now, let  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  be feature mappings such that

$$k_1(\mathbf{x}, \mathbf{x}') = \langle \boldsymbol{\alpha}(\mathbf{x}), \boldsymbol{\alpha}(\mathbf{x}') \rangle$$

$$k_2(\mathbf{x}, \mathbf{x}') = \langle \boldsymbol{\beta}(\mathbf{x}), \boldsymbol{\beta}(\mathbf{x}') \rangle$$

As a result,

$$\begin{aligned} k_1(\mathbf{x}, \mathbf{x}') k_2(\mathbf{x}, \mathbf{x}') &= \langle \boldsymbol{\alpha}(\mathbf{x}), \boldsymbol{\alpha}(\mathbf{x}') \rangle \langle \boldsymbol{\beta}(\mathbf{x}), \boldsymbol{\beta}(\mathbf{x}') \rangle \\ &= \boldsymbol{\alpha}(\mathbf{x})^T \boldsymbol{\alpha}(\mathbf{x}') \boldsymbol{\beta}^T(\mathbf{x}) \boldsymbol{\beta}(\mathbf{x}') \end{aligned}$$

$$\begin{aligned}
&= \left[ \sum_{i=1}^N \alpha_i(\mathbf{x}) \alpha_i(\mathbf{x}') \right] \left[ \sum_{j=1}^M \beta_j(\mathbf{x}) \beta_j(\mathbf{x}') \right] \\
&= \sum_{i=1}^N \sum_{j=1}^M \alpha_i(\mathbf{x}) \beta_j(\mathbf{x}) \alpha_i(\mathbf{x}') \beta_j(\mathbf{x}') \\
&= \sum_{i=1}^N \sum_{j=1}^M A_{ij}(\mathbf{x}) A_{ij}(\mathbf{x}') \\
&= \langle \mathbf{A}(\mathbf{x}), \mathbf{A}(\mathbf{x}') \rangle_{\mathbf{F}}
\end{aligned} \tag{*}$$

where  $\mathbf{A}$  is a matrix with

$$A_{ij}(\mathbf{x}) = \alpha_i(\mathbf{x}) \beta_j(\mathbf{x})$$

and  $\langle \cdot, \cdot \rangle_{\mathbf{F}}$  is the Frobenius inner product. Since the product kernel can be rewritten as a valid inner product in the feature space defined by the feature mapping  $\mathbf{A}(\mathbf{x})$ , the new kernel

$$k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}') k_2(\mathbf{x}, \mathbf{x}') \tag{6.18}$$

is valid. Note that we can continue differently from (\*), so

$$k_1(\mathbf{x}, \mathbf{x}') k_2(\mathbf{x}, \mathbf{x}') = \sum_{k=1}^K \phi_k(\mathbf{x}) \phi_k(\mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle$$

where  $K = NM$  and

$$\phi_k(\mathbf{x}) = \alpha_{((k-1) \oslash N) + 1}(\mathbf{x}) \beta_{((k-1) \odot N) + 1}(\mathbf{x})$$

where  $\oslash$  and  $\odot$  denote integer division and remainder, respectively. □

## Exercise 6.8 ★

Verify the results (6.19) and (6.20) for constructing valid kernels.

*Proof.* Let  $\psi$  be a feature mapping such that

$$k_3(\mathbf{x}, \mathbf{x}') = \langle \psi(\mathbf{x}), \psi(\mathbf{x}') \rangle$$

Then,

$$\begin{aligned}
k_3(\phi(\mathbf{x}), \phi(\mathbf{x}')) &= \langle \psi(\phi(\mathbf{x})), \psi(\phi(\mathbf{x}')) \rangle \\
&= \langle (\psi \circ \phi)(\mathbf{x}), (\psi \circ \phi)(\mathbf{x}') \rangle \\
&= \langle \gamma(\mathbf{x}), \gamma(\mathbf{x}') \rangle
\end{aligned}$$

where  $\phi$  is a function from  $\mathbf{x}$  to  $\mathbb{R}^M$  and  $\gamma = \psi \circ \phi$ . Therefore, the kernel

$$k(\mathbf{x}, \mathbf{x}') = k_3(\phi(\mathbf{x}), \phi(\mathbf{x}')) \tag{6.19}$$

is valid. For the second part, since  $\mathbf{A}$  is a symmetric, positive semidefinite matrix, one can use the Cholesky decomposition to obtain a matrix  $\mathbf{L}$  such that

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T$$

As a result, one can show that

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T \mathbf{L} \mathbf{L}^T \mathbf{x} = (\mathbf{L}^T \mathbf{x})^T (\mathbf{L}^T \mathbf{x}) = \langle \boldsymbol{\zeta}(\mathbf{x}), \boldsymbol{\zeta}(\mathbf{x}') \rangle$$

where  $\boldsymbol{\zeta}(\mathbf{x}) = \mathbf{L}^T \mathbf{x}$ . Hence, the kernel

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{A} \mathbf{x} \quad (6.20)$$

is valid.  $\square$

## Exercise 6.9 $\star$

Verify the results (6.21) and (6.22) for constructing valid kernels.

*Proof.* Let  $\phi_a$  and  $\phi_b$  be feature mappings so that

$$k_a(\mathbf{x}, \mathbf{x}') = \langle \phi_a(\mathbf{x}), \phi_a(\mathbf{x}') \rangle$$

$$k_b(\mathbf{x}, \mathbf{x}') = \langle \phi_b(\mathbf{x}), \phi_b(\mathbf{x}') \rangle$$

Therefore, since the inner product becomes a bilinear form on  $\mathbb{R}$ ,

$$\begin{aligned} k_a(\mathbf{x}_a, \mathbf{x}'_a) + k_b(\mathbf{x}_b, \mathbf{x}'_b) &= \langle \phi_a(\mathbf{x}_a), \phi_a(\mathbf{x}'_a) \rangle + \langle \phi_b(\mathbf{x}_b), \phi_b(\mathbf{x}'_b) \rangle \\ &= \langle (\phi_a(\mathbf{x}_a), \phi_a(\mathbf{x}'_a)), (\phi_b(\mathbf{x}_b), \phi_b(\mathbf{x}'_b)) \rangle \\ &= \langle \boldsymbol{\phi}(\mathbf{x}), \boldsymbol{\phi}(\mathbf{x}') \rangle \end{aligned}$$

where

$$\boldsymbol{\phi}(\mathbf{x}) = \begin{bmatrix} \phi_a(\mathbf{x}_a) \\ \phi_b(\mathbf{x}_b) \end{bmatrix}$$

Hence, the kernel

$$k(\mathbf{x}, \mathbf{x}') = k_a(\mathbf{x}_a, \mathbf{x}'_a) + k_b(\mathbf{x}_b, \mathbf{x}'_b) \quad (6.21)$$

is valid. The product identity is obtained similarly to what we do in Exercise 6.7. One has that

$$\begin{aligned} k_a(\mathbf{x}_a, \mathbf{x}'_a) k_b(\mathbf{x}_b, \mathbf{x}'_b) &= \langle \phi_a(\mathbf{x}_a), \phi_a(\mathbf{x}'_a) \rangle \langle \phi_b(\mathbf{x}_b), \phi_b(\mathbf{x}'_b) \rangle \\ &= \left[ \sum_{i=1}^{N_a} \phi_{ai}(\mathbf{x}_a) \phi_{ai}(\mathbf{x}'_a) \right] \left[ \sum_{j=1}^{N_b} \phi_{bj}(\mathbf{x}_b) \phi_{bj}(\mathbf{x}'_b) \right] \\ &= \sum_{i=1}^{N_a} \sum_{j=1}^{N_b} \phi_{ai}(\mathbf{x}_a) \phi_{bj}(\mathbf{x}_b) \phi_{ai}(\mathbf{x}'_a) \phi_{bj}(\mathbf{x}'_b) \\ &= \sum_{i=1}^{N_a} \sum_{j=1}^{N_b} A_{ij}(\mathbf{x}) A_{ij}(\mathbf{x}') \\ &= \langle \mathbf{A}(\mathbf{x}), \mathbf{A}(\mathbf{x}') \rangle_{\mathbf{F}} \end{aligned}$$

where  $\langle \cdot, \cdot \rangle_{\mathbf{F}}$  is the Frobenius inner product,  $\phi_{ai}(\mathbf{x})$  is the  $i$ -th element of  $\phi_a(\mathbf{x})$  and

$$A_{ij}(\mathbf{x}) = \phi_{ai}(\mathbf{x}_a) \phi_{bj}(\mathbf{x}_b)$$

Therefore, the new kernel

$$k(\mathbf{x}, \mathbf{x}') = k_a(\mathbf{x}_a, \mathbf{x}'_a) k_b(\mathbf{x}_b, \mathbf{x}'_b) \quad (6.22)$$

will also be valid.  $\square$

## Exercise 6.10 ★

Show that an excellent choice of kernel for learning a function  $f(\mathbf{x})$  is given by  $k(\mathbf{x}, \mathbf{x}') = f(\mathbf{x})f(\mathbf{x}')$  by showing that a linear learning machine-based on this kernel will always find a solution proportional to  $f(\mathbf{x})$ .

*Proof.* By substituting the kernel and (6.8) into (6.9), one has that

$$y(\mathbf{x}) = \mathbf{k}(\mathbf{x})^T (\mathbf{K} + \lambda \mathbf{I}_N)^{-1} \mathbf{t} = \mathbf{k}(\mathbf{x})^T \mathbf{a} = \sum_{n=1}^N k(\mathbf{x}, \mathbf{x}_n) a_n = f(\mathbf{x}) \left[ \sum_{n=1}^N f(\mathbf{x}_n) a_n \right]$$

which shows that the prediction function will always be proportional to  $f(\mathbf{x})$ .  $\square$

## Exercise 6.11 ★

By making use of the expansion (6.25), and then expanding the middle factor as a power series, show that the Gaussian kernel (6.23) can be expressed as the inner product of an infinite-dimensional feature vector.

*Proof.* We've seen in Section 6.2 that the Gaussian kernel can be expanded as

$$k(\mathbf{x}, \mathbf{x}') = \exp \left\{ -\frac{\|\mathbf{x}\|^2}{2\sigma^2} \right\} \exp \left\{ \frac{\langle \mathbf{x}, \mathbf{x}' \rangle}{\sigma^2} \right\} \exp \left\{ -\frac{\|\mathbf{x}'\|^2}{2\sigma^2} \right\} \quad (6.25)$$

In Exercise 6.7 we proved that if  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  are feature maps, there exists a feature map  $\boldsymbol{\psi}$  such that

$$\langle \boldsymbol{\alpha}(\mathbf{x}), \boldsymbol{\alpha}(\mathbf{x}') \rangle \langle \boldsymbol{\beta}(\mathbf{x}), \boldsymbol{\beta}(\mathbf{x}') \rangle = \langle \boldsymbol{\psi}(\mathbf{x}), \boldsymbol{\psi}(\mathbf{x}') \rangle$$

Therefore, one can prove using induction that there exists a feature map  $\boldsymbol{\zeta}$  such that for  $n \in \mathbb{N}$ ,

$$\langle \boldsymbol{\alpha}(\mathbf{x}), \boldsymbol{\alpha}(\mathbf{x}') \rangle^n = \langle \boldsymbol{\zeta}(\mathbf{x}), \boldsymbol{\zeta}(\mathbf{x}') \rangle$$

Now, using the definition of the exponential function for the middle term gives

$$\exp \left\{ \frac{\langle \mathbf{x}, \mathbf{x}' \rangle}{\sigma^2} \right\} = \sum_{i=0}^{\infty} \frac{1}{i! \sigma^{2i}} \langle \mathbf{x}, \mathbf{x}' \rangle^i = \sum_{i=0}^{\infty} \frac{1}{i! \sigma^{2i}} \langle \boldsymbol{\Psi}_i(\mathbf{x}), \boldsymbol{\Psi}_i(\mathbf{x}') \rangle = \sum_{i=0}^{\infty} \left\langle \frac{1}{\sigma} \sqrt{\frac{1}{i!}} \boldsymbol{\Psi}_i(\mathbf{x}), \frac{1}{\sigma} \sqrt{\frac{1}{i!}} \boldsymbol{\Psi}_i(\mathbf{x}') \right\rangle$$

where  $\boldsymbol{\Psi}_i$  are feature maps such that

$$\langle \mathbf{x}, \mathbf{x}' \rangle^i = \langle \boldsymbol{\Psi}_i(\mathbf{x}), \boldsymbol{\Psi}_i(\mathbf{x}') \rangle$$

Substituting this result back into (6.25) yields

$$\begin{aligned} k(\mathbf{x}, \mathbf{x}') &= \exp \left\{ -\frac{\|\mathbf{x}\|^2}{2\sigma^2} \right\} \exp \left\{ -\frac{\|\mathbf{x}'\|^2}{2\sigma^2} \right\} \sum_{i=0}^{\infty} \left\langle \frac{1}{\sigma} \sqrt{\frac{1}{i!}} \boldsymbol{\Psi}_i(\mathbf{x}), \frac{1}{\sigma} \sqrt{\frac{1}{i!}} \boldsymbol{\Psi}_i(\mathbf{x}') \right\rangle \\ &= \sum_{i=0}^{\infty} \exp \left\{ -\frac{\|\mathbf{x}\|^2}{2\sigma^2} \right\} \exp \left\{ -\frac{\|\mathbf{x}'\|^2}{2\sigma^2} \right\} \left\langle \frac{1}{\sigma} \sqrt{\frac{1}{i!}} \boldsymbol{\Psi}_i(\mathbf{x}), \frac{1}{\sigma} \sqrt{\frac{1}{i!}} \boldsymbol{\Psi}_i(\mathbf{x}') \right\rangle \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=0}^{\infty} \left\langle \frac{1}{\sigma} \sqrt{\frac{1}{i!}} \Psi_i(\mathbf{x}) \exp \left\{ -\frac{\|\mathbf{x}\|^2}{2\sigma^2} \right\}, \frac{1}{\sigma} \sqrt{\frac{1}{i!}} \Psi_i(\mathbf{x}') \exp \left\{ -\frac{\|\mathbf{x}'\|^2}{2\sigma^2} \right\} \right\rangle \\
&= \sum_{i=0}^{\infty} \phi_i(\mathbf{x}) \phi_i(\mathbf{x}') \\
&= \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle
\end{aligned}$$

where  $\phi(\mathbf{x})$  is a feature vector of infinite dimensionality with

$$\phi_i(\mathbf{x}) = \left\langle \frac{1}{\sigma} \sqrt{\frac{1}{i!}} \Psi_i(\mathbf{x}) \exp \left\{ -\frac{\|\mathbf{x}\|^2}{2\sigma^2} \right\} \right\rangle$$

□

## Exercise 6.12 ★★

Consider the space of all possible subsets  $A$  of a given fixed set  $D$ . Show that the kernel function (6.27) corresponds to an inner product in a feature space of dimensionality  $2^{|D|}$  defined by the mapping  $\phi(A)$  where  $A$  is a subset of  $D$  and the element  $\phi_U(\mathbf{A})$ , indexed by the subset  $U$ , is given by

$$\phi_U(A) = \begin{cases} 1, & \text{if } U \subseteq A \\ 0, & \text{otherwise} \end{cases} \quad (6.95)$$

Here  $U \subseteq A$  denotes that  $U$  is either a subset of  $A$  or is equal to  $A$ .

*Proof.* Using simple combinatorics, one can easily show that the number of subsets of a given fixed set  $D$  is given by  $2^{|D|}$ . Therefore,  $\phi(A)$  will be of dimensionality  $2^{|D|}$ . Since the element  $\phi_U(A)$  is 1 if  $U \subseteq A$  and 0 otherwise, the result of the inner product  $\langle \phi(A_1), \phi(A_2) \rangle$  will give the number of subsets of  $D$  contained by both  $A_1$  and  $A_2$ . However, since  $A_1, A_2 \subseteq D$  this can also be expressed by counting the number of subsets of  $A_1 \cap A_2$ . This is done by the kernel

$$k(A_1, A_2) = 2^{|A_1 \cap A_2|} \quad (6.27)$$

Hence, the kernel can be written as an inner product in the space defined by the mapping  $\phi(A)$  since

$$k(A_1, A_2) = 2^{|A_1 \cap A_2|} = \langle \phi(A_1), \phi(A_2) \rangle$$

□