Chapter 6 Kernel Methods

Yue Yu

6.1 Dual Representations

(6.3)

By setting the gradient of $J(\mathbf{w})$ with respect to $\mathbf{w}$ to $\mathbf{0}$, it is easy to see that

$$\begin{aligned}
\mathbf{w} &= \arg_{\mathbf{w}} \nabla_{\mathbf{w}} J(\mathbf{w}) = \mathbf{0} \\
&= \arg_{\mathbf{w}} \left(\sum_{n=1}^{N} \left\{ \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) - t_n \right\} + \lambda \mathbf{w} = \mathbf{0} \right) \\
&= -\frac{1}{\lambda} \sum_{n=1}^{N} \left\{ \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) - t_n \right\} \boldsymbol{\phi}(\mathbf{x}_n) \\
&= \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{a}
\end{aligned}$$

where

$$\mathbf{a} = -\frac{1}{\lambda}(\boldsymbol{\Phi}\mathbf{w} - \mathbf{t}).$$

(6.5)

$$\begin{aligned}
J(\mathbf{a}) &= \frac{1}{2} \sum_{n=1}^{N} \left\{ \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) - t_n \right\}^2 + \frac{\lambda}{2} \mathbf{w}^{\mathrm{T}} \mathbf{w} \\
&= \frac{1}{2} \sum_{n=1}^{N} \left\{ \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} \mathbf{w} - t_n \right\}^2 + \frac{\lambda}{2} \mathbf{w}^{\mathrm{T}} \mathbf{w} \\
&= \frac{1}{2} \sum_{n=1}^{N} \left\{ \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} \mathbf{w} - 2\mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) t_n + t_n^2 \right\} + \frac{\lambda}{2} \mathbf{w}^{\mathrm{T}} \mathbf{w} \\
&= \frac{1}{2} \mathbf{w}^{\mathrm{T}} \left(\sum_{n=1}^{N} \boldsymbol{\phi}(\mathbf{x}_n) \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} \right) \mathbf{w} - \mathbf{w}^{\mathrm{T}} \sum_{n=1}^{N} \boldsymbol{\phi}(\mathbf{x}_n) t_n + \frac{1}{2} \sum_{n=1}^{N} t_n^2 + \frac{\lambda}{2} \mathbf{w}^{\mathrm{T}} \mathbf{w} \\
&= \frac{1}{2} \mathbf{a}^{\mathrm{T}} \boldsymbol{\Phi} \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi} \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{a} - \mathbf{a}^{\mathrm{T}} \boldsymbol{\Phi} \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{t} + \frac{1}{2} \mathbf{t}^{\mathrm{T}} \mathbf{t} + \frac{\lambda}{2} \mathbf{a}^{\mathrm{T}} \mathbf{K} \mathbf{a}.
\end{aligned}$$

(6.8)

From the derivation of (6.3), we have

$$\begin{aligned}
\mathbf{a} &= -\frac{1}{\lambda}(\boldsymbol{\Phi}\mathbf{w} - \mathbf{t}) \\
&= -\frac{1}{\lambda}(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{a} - \mathbf{t}) \\
&= -\frac{1}{\lambda}(\mathbf{K}\mathbf{a} - \mathbf{t}).
\end{aligned}$$

Solving for $\mathbf{a}$, we obtain

$$\mathbf{a} = (\mathbf{K} + \lambda \mathbf{I}_N)^{-1} \mathbf{t}.$$

(6.9)

$$y(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\Phi}\boldsymbol{\phi}(\mathbf{x}) = (\boldsymbol{\Phi}\boldsymbol{\phi}(\mathbf{x}))^{\mathrm{T}}\mathbf{a} = \mathbf{k}(\mathbf{x})^{\mathrm{T}}(\mathbf{K} + \lambda\mathbf{I}_N)^{-1}\mathbf{t}.$$

6.2 Constructing Kernels

(6.13)

To verify (6.13) is a valid kernel, we see that for any vector $\mathbf{u} \in \mathbb{R}^N$,

$$\mathbf{u}^{\mathrm{T}}\mathbf{K}\mathbf{u} = \mathbf{u}^{\mathrm{T}}(c\mathbf{K}_1)\mathbf{u} = c(\mathbf{u}^{\mathrm{T}}\mathbf{K}_1\mathbf{u}) \geq 0,$$

which implies that $\mathbf{K} \succeq 0$.

(6.14)

To verify (6.14) is a valid kernel, we first notice that $k(\mathbf{x}, \mathbf{x}')$ is symmetric by

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

Also, we see that for any vector $\mathbf{u} \in \mathbb{R}^N$,

$$\begin{aligned}
\mathbf{u}^{\mathrm{T}}\mathbf{K}\mathbf{u} &= \sum_{i=1}^{N}\sum_{j=1}^{N} k(\mathbf{x}_i, \mathbf{x}_j)u_i u_j \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} k_1(\mathbf{x}_i, \mathbf{x}_j)(f(\mathbf{x}_i)u_i)(f(\mathbf{x}_j)u_j) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} k_1(\mathbf{x}_i, \mathbf{x}_j)v_i v_j \\
&\geq 0
\end{aligned}$$

where we denote $v_i = f(\mathbf{x}_i)u_i$, implying $\mathbf{K} \succeq 0$. Therefore, (6.14) is a valid kernel.

(6.15)

To verify (6.15) is a valid kernel, let $q(k_1(\mathbf{x}, \mathbf{x}')) = \sum_m c_m k_1(\mathbf{x}, \mathbf{x}')^m$ where $c_m \geq 0$. By repeatedly applying (6.18), we see that $k_1(\mathbf{x}, \mathbf{x}')^m$ is a valid kernel. Then, according to (6.13), $c_m k_1(\mathbf{x}, \mathbf{x}')^m$ is also a valid kernel. Finally, using (6.17), we can conclude that $q(k_1(\mathbf{x}, \mathbf{x}')) = \sum_m c_m k_1(\mathbf{x}, \mathbf{x}')^m$ is a valid kernel.

(6.16)

To verify (6.16) is a valid kernel, we first notice that $k(\mathbf{x}, \mathbf{x}')$ is symmetric by

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

Also, we see that for any vector $\mathbf{u} \in \mathbb{R}^N$,

$$\begin{aligned}
\mathbf{u}^{\mathrm{T}}\mathbf{K}\mathbf{u} &= \sum_{i=1}^{N}\sum_{j=1}^{N} k(\mathbf{x}_i, \mathbf{x}_j)u_i u_j \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} \exp(k_1(\mathbf{x}_i, \mathbf{x}_j))u_i u_j.
\end{aligned}$$

Applying the Maclaurin series for the exponential term, we have

$$\mathbf{u}^{\mathrm{T}}\mathbf{K}\mathbf{u} = \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sum_{n=0}^{\infty}\frac{k_1(\mathbf{x}_i,\mathbf{x}_j)^n}{n!}\right)u_i u_j \geq 0$$

where we made use of (6.15), implying $\mathbf{K} \succeq 0$. Therefore, (6.16) is a valid kernel.

(6.17)

To verify (6.17) is a valid kernel, we first notice that $k(\mathbf{x},\mathbf{x}')$ is symmetric by

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

Also, we see that for any vector $\mathbf{u} \in \mathbb{R}^N$,

$$\begin{aligned}
\mathbf{u}^{\mathrm{T}}\mathbf{K}\mathbf{u} &= \sum_{i=1}^{N}\sum_{j=1}^{N} k(\mathbf{x}_i,\mathbf{x}_j)u_i u_j \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}(k_1(\mathbf{x}_i,\mathbf{x}_j) + k_2(\mathbf{x}_i,\mathbf{x}_j))u_i u_j \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} k_1(\mathbf{x}_i,\mathbf{x}_j)u_i u_j + \sum_{i=1}^{N}\sum_{j=1}^{N} k_2(\mathbf{x}_i,\mathbf{x}_j)u_i u_j \\
&\geq 0,
\end{aligned}$$

which implies that $\mathbf{K} \succeq 0$. Therefore, (6.17) is a valid kernel.

(6.18)

To verify (6.18) is a valid kernel, we first notice that $k(\mathbf{x},\mathbf{x}')$ is symmetric by

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

Also, we see that for any vector $\mathbf{u} \in \mathbb{R}^N$,

$$\begin{aligned}
\mathbf{u}^{\mathrm{T}}\mathbf{K}\mathbf{u} &= \sum_{i=1}^{N}\sum_{j=1}^{N} k(\mathbf{x}_i,\mathbf{x}_j)u_i u_j \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} k_1(\mathbf{x}_i,\mathbf{x}_j)k_2(\mathbf{x}_i,\mathbf{x}_j)u_i u_j \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} \boldsymbol{\phi}_1(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\phi}_1(\mathbf{x}_j)\boldsymbol{\phi}_2(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\phi}_2(\mathbf{x}_j)u_i u_j \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sum_{l=1}^{M}\phi_{1l}(\mathbf{x}_i)\phi_{1l}(\mathbf{x}_j)\right)\left(\sum_{m=1}^{M}\phi_{2m}(\mathbf{x}_i)\phi_{2m}(\mathbf{x}_j)\right)u_i u_j \\
&= \sum_{l=1}^{M}\sum_{m=1}^{M}\left(\sum_{i=1}^{N}\phi_{1l}(\mathbf{x}_i)\phi_{2m}(\mathbf{x}_i)u_i\right)^2 \\
&\geq 0,
\end{aligned}$$

which implies that $\mathbf{K} \succeq 0$. Therefore, (6.18) is a valid kernel.

(6.20)

To verify (6.20) is a valid kernel, recall that any symmetric matrix $\mathbf{S}$ can be diagonalized by $\mathbf{S} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{\mathrm{T}}$ where the columns of $\mathbf{Q}$ are orthogonal eigenvectors. Hence,

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}' = \mathbf{x}^{\mathrm{T}}\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{\mathrm{T}}\mathbf{x}' = (\mathbf{Q}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}\boldsymbol{\Lambda}(\mathbf{Q}^{\mathrm{T}}\mathbf{x}').$$

Denoting $\mathbf{v} = \mathbf{Q}^{\mathrm{T}}\mathbf{x}$ and $\mathbf{v}' = \mathbf{Q}^{\mathrm{T}}\mathbf{x}'$,

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{v}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{v}' = \sum_{i=1}^{N} \Lambda_{ii} v_i v_i' = \sum_{i=1}^{N} (\sqrt{\Lambda_{ii}} v_i)(\sqrt{\Lambda_{ii}} v_j) = \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}') = k'(\boldsymbol{\phi}(\mathbf{x}), \boldsymbol{\phi}(\mathbf{x}'))$$

where k' is the linear kernel, and the square roots exist because the eigenvalues for a positive semidefinite matrix are non-negative. Therefore, according to (6.19), $k(\mathbf{x}, \mathbf{x}')$ is a valid kernel.

(6.21)

$$\begin{aligned}
k(\mathbf{x}, \mathbf{x}') &= k_a(\mathbf{x}_a, \mathbf{x}_a') + k_b(\mathbf{x}_b, \mathbf{x}_b') \\
&= \sum_{s=1}^{S} \phi_{as}(\mathbf{x}_a)\phi_{as}(\mathbf{x}_a') + \sum_{t=1}^{T} \phi_{bt}(\mathbf{x}_b)\phi_{bt}(\mathbf{x}_b') \\
&= \sum_{i=1}^{S+T} \phi_i(\mathbf{x})\phi_i(\mathbf{x}'),
\end{aligned}$$

which is a valid kernel.

(6.22) – To be updated

$$\begin{aligned}
k(\mathbf{x}, \mathbf{x}') &= k_a(\mathbf{x}_a, \mathbf{x}_a')k_b(\mathbf{x}_b, \mathbf{x}_b') \\
&= \sum_{s=1}^{S} \phi_{as}(\mathbf{x}_a)\phi_{as}(\mathbf{x}_a') \sum_{t=1}^{T} \phi_{bt}(\mathbf{x}_b)\phi_{bt}(\mathbf{x}_b') \\
&= \sum_{s=1}^{S}\sum_{t=1}^{T} \big\{\phi_{as}(\mathbf{x}_a)\phi_{bt}(\mathbf{x}_b)\big\}\big\{\phi_{as}(\mathbf{x}_a')\phi_{bt}(\mathbf{x}_b')\big\}
\end{aligned}$$