Part. 2, Questions (50%):

1. (10%) Given a valid kernel $k_1(x, x')$, prove that the following proposed functions are or are not valid kernels.
   a. $k(x, x') = (k_1(x, x'))^2 + (k_1(x, x') + 1)^2$
   b. $k(x, x') = (k_1(x, x'))^2 + \exp(\|x\|^2) * \exp(\|x'\|^2)$

2. (10%) Show that the kernel matrix $\mathbf{K} = [k(\mathbf{x}_n, \mathbf{x}_m)]_{nm}$ should be positive semidefinite is the necessary and sufficient condition for to be $k(\mathbf{x}, \mathbf{x}')$ a valid kernel.

3. (10%) Consider the dual formulation of the least-squares linear regression problem given on page 6 in the ppt of Kernel Methods. Show that the solution for the components $\mathbf{a}_n$ of the vector $\mathbf{a}$ can be expressed as a linear combination of the elements of the vector $\varphi(\mathrm{x}_n)$. Denoting these coefficients by the vector w, show that the dual of the dual formulation is given by the original representation in terms of the parameter vector w.

4. (10%) Prove that the Gaussian kernel defined by (eq 1) is valid and show the function $\varphi(\mathrm{x})$, where $\mathrm{x} \in \mathbf{R}^1$.

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\|\mathbf{x} - \mathbf{x}'\|^2 / 2\sigma^2\right) = \phi(x)^{\mathrm{T}} \phi(x')$$

(eq1)

5. (10%) Consider the optimization problem

$$\text{minimize } (x - 2)^2$$
$$\text{subject to } (x+3)(x-1) \leqq 2$$

State the dual problem.

1. (10%) Given a valid kernel $k_1(x, x')$, prove that the following proposed functions are or are not valid kernels.
   a. $k(x, x') = (k_1(x, x'))^2 + (k_1(x, x') + 1)^2$
   b. $k(x, x') = (k_1(x, x'))^2 + \exp(\|x\|^2) * \exp(\|x'\|^2)$

1. (a) ① $k_1(x, x)' + 1$ is a valid kernel function

For any positive constant $\alpha \geq 0$, the function

$K(x, x') = K_1(x, x') + \alpha$ is a valid kernal function

Proof: Let ϕ_1 denote a feature map of K_1. Then, using the feature map $\phi: x \mapsto [\phi_1(x), \sqrt{\alpha}]^T$, we have

$K(x, x') = \langle \phi(x), \phi(x') \rangle = \langle \phi_1(x), \phi_1(x') \rangle + \alpha = K_1(x, x') + \alpha$

Let $\alpha = 1$, ① is proved

② $[k_1(x, x')]^2$ 、$[k_1(x, x') + 1]^2$ are valid kernel function

From equation 6.18: $k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}')k_2(\mathbf{x}, \mathbf{x}')$ is a valid kernal function if $k_1(x, x')$、$k_2(x, x')$ are valid kernel function

let $k_1(x, x') = k_2(x, x') = k_1(x, x')$, $k_1(x, x') = k_2(x, x') = k_1(x, x') + 1$

② is proved

③ $k(x, x') = [k_1(x, x')]^2 + [k_1(x, x') + 1]^2$ is a valid kernel function

From equation 6.17: $k(\mathbf{x}, \mathbf{x}') = k_1(\mathbf{x}, \mathbf{x}') + k_2(\mathbf{x}, \mathbf{x}')$ is a valid kernel function if $k_1(x, x')$、$k_2(x, x')$ are valid kernel function

Let $k_1(x, x') = [k_1(x, x')]^2$, $k_2(x, x') = [k_1(x, x') + 1]$, ③ is proved

#

(b)

$$\exp(\|x\|^2)\exp(\|x'\|^2) = \exp(\|x\|^2 + \|x'\|^2)$$

$$= \exp(x^T x + (x')^T x') = \exp(x^T x + x^T x' + (x')^T x' - x^T x')$$

$$= \underbrace{\exp(x^T x)\exp(x^T x')\exp((x')^T x')}\,\underbrace{\exp(-x^T x')}$$

From eq. 6.14 $k(\mathbf{x},\mathbf{x}') = f(\mathbf{x})k_1(\mathbf{x},\mathbf{x}')f(\mathbf{x}')$

Let $f(x) = \exp(x^T x)$, $k_1(x,x') = \exp(x^T x')$, $f(x') = \exp((x')^T x')$

$k_e(x,x') = x^T x'$ is a valid kernel $\because$ The feature space exists $\phi(x) = x$

$\Rightarrow$ $k_1(x,x') = \exp(x^T x')$ is a valid kernel by eq. 6.16

So does $k_2(x,x') = \exp(-x^T x')$

From eq. 6.18 $k(\mathbf{x},\mathbf{x}') = k_1(\mathbf{x},\mathbf{x}')k_2(\mathbf{x},\mathbf{x}')$ is a valid kernel if $k_1(x,x')$, $k_2(x,x')$ are valid kernels

Let $k_1(x,x') = \exp(x^T x)\exp(x^T x')\exp((x')^T x')$, $k_2(x,x') = \exp(-x^T x')$

$\Rightarrow$ $k_2(x,x') = \exp(\|x\|^2)\exp(\|x'\|^2)$ is valid

From eq. 6.17 $k(\mathbf{x},\mathbf{x}') = k_1(\mathbf{x},\mathbf{x}') + k_2(\mathbf{x},\mathbf{x}')$ is a valid kernal if $k_1(x,x')$, $k_2(x,x')$ are valid kernels

Let $k_1(x,x') = [k_1(x,x')]^2$ which is a valid kernel proved by 1. (a), $k_2(x,x') = \exp(\|x\|^2)\exp(\|x'\|^2)$

$\Rightarrow$ $k(x,x') = [k_1(x,x')]^2 + \exp(\|x\|^2)\exp(\|x'\|^2)$ is a valid kernel

※

2. (10%) Show that the kernel matrix s $\mathbf{K} = [k(\mathbf{x}_n, \mathbf{x}_m)]_{nm}$ hould be positive sem idefinite is the necessary and sufficient condition for to be $k(\mathbf{x}, \mathbf{x}')$ a valid kernel.

$k(x, x')$ is a valid kernal

$\Rightarrow$ $K = [k(x_n, x_m)]_{nm}$ is positive semidefinite kernel matrix

Proof: Assume k is a valid kernel, then

$$K_{ij} = K(x^{(i)}, x^{(j)}) = \phi(x(i))^T \phi(x(j))$$

$$= \phi(x(j))^T \phi(x(i)) = K(x^{(j)}, x^{(i)}) = K_{ji}$$

$\Rightarrow$ K is a symmetric matrix ... ①

Let $\phi_k(x)$ denote the k th element of $\phi(x)$, for any vector z:

$$z^T K z = \sum_i \sum_j z_i K_{ij} z_j$$

$$= \sum_i \sum_j z_i \phi(x^{(i)})^T \phi(x^{(j)}) z_j$$

$$= \sum_i \sum_j z_i \sum_k \phi_k(x^{(i)}) \phi_k(x^{(j)}) z_j$$

$$= \sum_k \sum_i \sum_j z_i \phi_k(x^{(i)}) \phi_k(x^{(j)}) z_j$$

$$= \sum_k \left[\sum_i z_i \phi_k(x^{(i)}) \right] \left[\sum_j z_j \phi_k(x^{(j)}) \right]$$

$$= \sum_k \left[\sum_i z_i \phi_k(x^{(i)}) \right]^2 \geq 0$$

$\Rightarrow$ All eigenvalues are non-negative ... ②

By ①, ② $\Rightarrow$ $K = [k(x_n, x_m)]_{nm}$ is positive semidefinite kernel matrix

$K = [k(x_n, x_m)]_{nm}$ is positive semidefinite kernel matrix

$\Rightarrow k(x, x')$ is a valid kernel

Proof:

Assume $K = [k(x^{(i)}, x^{(j)})]_{ij}$ is positive semidefinite kernel matrix

K is symmetric. Thus, we have $K = V \Lambda V^T$, where V is an orthonormal matrix V_t and the diagonal matrix Λ contains the eigenvalues λ_t of K. Since K is positive semidefinite, all eigenvalues λ_t are non-negative.

Consider the feature map: $\phi: x_i \mapsto (\sqrt{\lambda_t}\, V_{ti})_{t=1}^{n} \in \mathbb{R}^n$

$$\Rightarrow \phi(x_i)^T \phi(x_j) = \sum_{t=1}^{n} \lambda_t V_{ti} V_{tj} = (V \Lambda V^T)_{ij} = K_{ij} = k(x_i, x_j)$$

There exists a space where the output space is equal to the inner product of two data points

※

3. (10%) Consider the dual formulation of the least-squares linear regression pro blem given on page 6 in the ppt of Kernel Methods. Show that the solution fo r the components $\mathbf{a}_n$ of the vector $\mathbf{a}$ can be expressed as a linear combination o f the elements of the vector $\boldsymbol{\varphi}(\mathbf{x}_n)$. Denoting these coefficients by the vector $\mathbf{w}$, show that the dual of the dual formulation is given by the original represen tation in terms of the parameter vector $\mathbf{w}$.

We first note that $J(a)$ depends on a only through the form Ka.
Since typically the number N of data points is greater than the number M of basis functions, the matrix $K = \Phi\Phi^T$ will be rank deficient.
There will be M eigenvectors of K having non-zero eigenvalues, and $N-M$ eigenvalues with eigenvalue zero. We can then decompose $a = a_{\parallel} + a_{\perp}$ where $a_{\parallel}^T a_{\perp} = 0$ and $K a_{\perp} = 0$.
Thus the value of $a_{\perp}$ is not determined by $J(a)$. We can remove the ambiguity by setting $a_{\perp} = 0$, or equivalently by adding a regularizer term $\frac{\varepsilon}{2} a_{\perp}^T a_{\perp}$ to $J(a)$ where ε is a small positive constant. Then $a = a_{\parallel}$ where $a_{\parallel}$ lies in the span of $K = \Phi\Phi^T$ and hence can be writen as a linear combination of the columns of Φ, so that in component notation $a_n = \sum_{i=1}^{M} u_i \phi_i(x_n)$ or equivalently in vector notation $a = \Phi u$ ... ①

Substituting ① into equation 6.7: $J(a) = \frac{1}{2} a^T K K a - a^T K t + \frac{1}{2} t^T t + \frac{\lambda}{2} a^T K a$
we obtain:

$$J(u) = \frac{1}{2}(K\Phi u - t)^T (K\Phi u - t) + \frac{\lambda}{2} u^T K \Phi u$$

$$= \frac{1}{2}(\Phi\Phi^T\Phi u - t)^T(\Phi\Phi^T\Phi u - t) + \frac{\lambda}{2} u^T \Phi^T \Phi\Phi^T \Phi u \quad \ldots ②$$

Since the matrix $\Phi^T\Phi$ has full rank we can define an equivalent parametrization given by $w = \Phi^T \Phi u$ and substitutin this into ②:

$$J(w) = \frac{1}{2}(\Phi w - t)^T(\Phi w - t) + \frac{\lambda}{2} w^T w = \frac{1}{2}\sum_{n=1}^{N}\{w^T\phi(x_n) - t_n\}^2 + \frac{\lambda}{2} w^T w$$

We recover the original regularized error function 6.2.

※

4. (10%) Prove that the Gaussian kernel defined by (eq 1) is valid and show the function φ (x), where x $\in$ **R**[1].

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\|\mathbf{x}-\mathbf{x}'\|^2/2\sigma^2\right) = \phi(x)^{\mathrm{T}}\phi(x')$$

(eq1)

① $\|x-x'\|^2 = x^T x + (x')^T x' - 2x^T x'$

$\Rightarrow k(x,x') = \exp(-x^T x/2\sigma^2)\exp(x^T x'/\sigma^2)\exp(-(x')^T x'/2\sigma^2)$

We make use of 6.14 $k(\mathbf{x}, \mathbf{x}') = f(\mathbf{x})k_1(\mathbf{x}, \mathbf{x}')f(\mathbf{x}')$ and 6.16 $k(\mathbf{x}, \mathbf{x}') = \exp(k_1(\mathbf{x}, \mathbf{x}'))$, together with the validity of the linear kernel $k(x,x') = x^T x'$.

② Considering only 1 dimension of input vector $X = x$

$k(x,x') = \exp(-(x^2)/2\sigma^2)\exp(xx'/\sigma^2)\exp(-(x')^2/2\sigma^2)$

$\exp(xx'/\sigma^2) = \sum_{i=0}^{\infty}\frac{(xx'/\sigma^2)^i}{i!}$ based on Taylor expansion

$$\begin{aligned}
k(x,x') &= \exp(-(x^2)/2\sigma^2)\exp(-(x')^2/2\sigma^2)\sum_{i=0}^{\infty}\frac{(xx'/\sigma^2)^i}{i!} \\
&= \sum_{i=0}^{\infty}\exp(-(x^2)/2\sigma^2)\exp(-(x')^2/2\sigma^2)\frac{(x/\sigma)^i}{\sqrt{i!}}\cdot\frac{(x'/\sigma)^i}{\sqrt{i!}} \\
&= \sum_{i=0}^{\infty}\left[\exp(-(x^2)/2\sigma^2)\frac{(x/\sigma)^i}{\sqrt{i!}}\right]\left[\exp(-(x')^2/2\sigma^2)\frac{(x'/\sigma)^i}{\sqrt{i!}}\right]
\end{aligned}$$

with infinite dimensional

$$\phi(x) = \exp(-(x^2)/2\sigma^2)\cdot\left(1, \frac{x}{\sigma}, \frac{1}{\sqrt{2}}\left(\frac{x}{\sigma}\right)^2, \frac{1}{\sqrt{6}}\left(\frac{x}{\sigma}\right)^3, \ldots, \frac{1}{\sqrt{n!}}\left(\frac{x}{\sigma}\right)^n, \ldots\right)$$

※

5. (10%) Consider the optimization problem

minimize $(x - 2)^2$

subject to $(x+3)(x-1) \leqq 2$

State the dual problem.

$f_0(x) = (x-2)^2 \quad f_1(x) = (x+3)(x-1) - 2$

$L(x, \lambda) = f_0 + \lambda f_1 = (x-2)^2 + \lambda\left[(x+3)(x-1) - 2\right]$

$= (1+\lambda)x^2 + (-4+2\lambda)x + (4-5\lambda)$

$$\frac{\partial L(x,\lambda)}{\partial x} = 2(1+\lambda)x + (-4+2\lambda) = 0$$

$$\Rightarrow \quad x = \frac{4-2\lambda}{2(1+\lambda)} = \frac{2-\lambda}{1+\lambda}$$

dual representation

$\lambda \geq 0$

#