Machine Learning 1

Homework 8: Kernel Method

Student: Nguyễn Thị Minh Ngọc - ID: 11219280

1 Problem 1

Duel Representation

Solution. We consider a linear regression model whose parameters are determined by minimizing a regularized sum-of-squares error function given:

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

where $\lambda \geq 0$
Taking the derivative of J(w) with respect to w equal to zero:

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \sum_{n=1}^{N}\{\mathbf{w}^T\phi(x_n) - t_n\}\phi(x_n) + \lambda\mathbf{w} = 0$$

then:

$$\mathbf{w} = \frac{-1}{\lambda}\sum_{n=1}^{N}\{\mathbf{w}^T\phi(x_n) - t_n\}\phi(x_n)$$

Let $a_n = \frac{-1}{\lambda}\mathbf{w}^T\phi(x_n) - t_n$ $\Phi = [\phi(x_1)\ \phi(x_2)\ ...\ \phi(x_n)]^T$, then:

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \frac{-1}{\lambda}\begin{bmatrix} \mathbf{w}^T\phi(x_1) - t_1 \\ \mathbf{w}^T\phi(x_2) - t_2 \\ \vdots \\ \mathbf{w}^T\phi(x_n) - t_n \end{bmatrix}$$

and

$$\mathbf{w} = \Phi^T a$$

We have:

$$\begin{aligned} J(\mathbf{w}) &= \frac{1}{2}\sum_{n=1}^{N}\{\mathbf{w}^T\phi(x_n) - t_n\}^2 + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \\ &= \frac{1}{2}\|\Phi\mathbf{w} - t\|_2^2 + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \\ &= \frac{1}{2}(\Phi\mathbf{w} - t)^T(\Phi\mathbf{w} - t) + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \\ &= \frac{1}{2}(\mathbf{w}^T\Phi - t^T)(\Phi\mathbf{w} - t) + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \\ &= \frac{1}{2}(\mathbf{w}^T\Phi^T\Phi\mathbf{w} - \mathbf{w}^T\Phi t - t^T\Phi\mathbf{w} + t^Tt) + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \\ &= \frac{1}{2}\mathbf{w}^T\Phi^T\Phi\mathbf{w} - t^T\Phi\mathbf{w} + \frac{1}{2}t^Tt + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \end{aligned}$$

Substitutes $\mathbf{w} = \Phi^T a$:

$$J(\mathbf{w}) = \frac{1}{2} a^T \Phi \Phi^T \Phi \Phi^T a - t^T \Phi \Phi^T a + \frac{1}{2} t^T t + \frac{\lambda}{2} a^T \Phi \Phi^T a$$

We now define $\mathbf{K} = \Phi \Phi^T \in \mathbb{R}^{nxn}$, which is an N×N symmetric matrix with elements:

$$\mathbf{K}_{nm} = \phi(x_n)^T \phi(x_m) = k(x_n, x_m)$$

Substitutes $\mathbf{K} = \Phi \Phi^T$:

$$J(\mathbf{w}) = \frac{1}{2} a^T \mathbf{K} \mathbf{K} a - t^T \mathbf{K} a + \frac{1}{2} t^T t + \frac{\lambda}{2} a^T \mathbf{K} a$$

Taking the derivative of J(w) with respect to a equal to zero:

$$\frac{\partial J(\mathbf{w})}{\partial a} = \mathbf{K}\mathbf{K}a - \mathbf{K}t + \lambda \mathbf{K} a = 0$$

Hence:

$$a = (K + \lambda I_N)^{-1} t$$

2 Problem 2

Prove that the results (6.13) and (6.14) are valid kernels.

Solution. To prove that the results (6.13) and (6.14) are valid kernels, we need to prove that they correspond to a scalar product in some (perhaps infinite dimensional) feature space.

If $k_1(x, x')$ is a valid kernel then there must exist a feature vector $\phi(x)$ such that:

$$k_1(x, x') = \phi(x)^T \phi(x')$$

(6.13) We have:

$$\begin{aligned} k(x, x') &= c k_1(x, x') \\ &= \sqrt{c}\phi(x)^T \sqrt{c}\phi(x') \\ &= \varphi(x)^T \varphi(x') \end{aligned}$$

where $\varphi(x) = \frac{c}{2}\phi(x)$ and $c > 0$
Hence $k(x, x') = ck_1(x, x')$ (6.13) is a valid kernel.

(6.14) We have:

$$\begin{aligned} k(x, x') &= f(x) k_1(x, x') f(x') \\ &= f(x)\phi(x)^T \phi(x') f(x') \\ &= \varphi(x)^T \varphi(x') \end{aligned}$$

where $\varphi(x) = f(x)\phi(x)^T$
Hence $k(x, x') = f(x)k_1(x, x')f(x')$ (6.14) is a valid kernel.

3 Problem 3

Dataset $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} -3 & 4 \\ 1 & 0 \end{bmatrix}$; $\phi(x) = [x_1, x_2, \|x\||]^T$. Calculate the Kernel Matrix K.
Solution.

$$x_1 = (-3, 4) \Rightarrow \phi(x_1) = [-3\ 4\ \sqrt{(-3)^2 + 4^2}]^T = [-3\ 1\ 5]^T$$
$$x_2 = (1, 0) \Rightarrow \phi(x_2) = [1\ 0\ \sqrt{1^2 + 0^2}]^T = [1\ 0\ 1]^T$$

We have:

$$\Phi = \begin{bmatrix} \phi(x_1) \\ \phi(x_2) \end{bmatrix} = \begin{bmatrix} -3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix}$$

then:

$$K = \Phi\Phi^T = \begin{bmatrix} -3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} -3 & 1 \\ 4 & 0 \\ 5 & 1 \end{bmatrix} = \begin{bmatrix} 50 & 2 \\ 2 & 2 \end{bmatrix}$$

Hence $K = \begin{bmatrix} 50 & 2 \\ 2 & 2 \end{bmatrix}$