

Machine Learning 1

Homework Week 8 - Kernel Method

Tran Hai Nam - 11219279 - DSEB

1 Problem 1: Dual Representation

L2 Loss regularization

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

where $\lambda > 0$

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

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

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

$$w = -\frac{1}{2} \sum_{i=1}^N (w^T \phi(x_n) - t_n) \cdot \phi(x_n)$$

$$= \sum_{n=1}^N a_n \phi(x_n) = \phi^\top \cdot a$$

$$\phi = \begin{bmatrix} \phi(x_1) \\ \phi(x_2) \\ \vdots \\ \phi(x_n) \end{bmatrix}$$

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$a_n = \frac{-1}{\lambda} (w^T \phi(x_n) - t_n)$$

→ Substitute $w = \phi^\top a$

$$\begin{aligned}
L &= \frac{1}{2} \sum_{l=1}^n (w^\top \phi(x_n) t_n)^2 + \frac{\lambda}{2} \|w\|_2^2 \\
&= \frac{1}{2} (w^\top \phi - t)^\top \cdot (w^\top \phi - t) + \frac{\lambda}{2} w^\top w \\
&= \frac{1}{2} (\phi^\top \cdot w - t^\top) (w^\top \phi - t) + \frac{\lambda}{2} w^\top w \\
&= \frac{1}{2} (\phi^\top w \cdot w^\top \phi - \phi^\top w \cdot t - t^\top \cdot w^\top \cdot \phi + t^\top \cdot t) + \frac{\lambda}{2} w^\top w \\
&= \frac{1}{2} (w^\top \cdot \phi^\top \cdot \phi \cdot w - 2w^\top \cdot \phi^\top \cdot t + t^\top \cdot t) + \frac{\lambda}{2} w^\top w \\
&= \frac{1}{2} a^\top \cdot \phi \cdot \phi^\top \cdot \phi \cdot \phi^\top \cdot a - a^\top \cdot \phi \cdot \phi^\top \cdot t + \frac{1}{2} t^\top \cdot t + \frac{\lambda}{2} a^\top \phi \cdot \phi^\top a
\end{aligned}$$

Let Kernel matrix $K = \phi\phi^T$. K is symmetric (a) and positive semi-definite (PSD) (b):

- $K^T = (\phi\phi^T)^T = (\phi^T)^T \phi^T = \phi\phi^T = K$ (a)
- $(Kx, x) \geq 0 \forall x \in R^n$ since:
 $(Kx, x) = x^T Kx = x^T \phi\phi^T x = (\phi^T x)^T (\phi^T x) = \|\phi^T x\|_2^2 \geq 0$

$$\begin{aligned}
\rightarrow J(a) &= \frac{1}{2} a^T K K a - a^T K t + \frac{1}{2} t^T t + \frac{\lambda}{2} a^T K a \\
\frac{\partial J}{\partial a} &= K K a - K t + \lambda K a = 0 \\
K(Ka - t + \lambda a) &= 0
\end{aligned}$$

Since K is PSD:

$$\begin{aligned}
Ka - t + \lambda a &= 0 \\
(K + \lambda I)a &= t \\
a &= (K + \lambda I)^{-1} t
\end{aligned}$$

2 Problem 2: Valid Kernels

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

Since $k_1(x, x')$ is a valid kernel → exist an feature vector $\phi(x)$ such that $k_1(x, x') = \phi(x)^T \phi(x')$.

Let $u(x) = \sqrt{c}k_1(x, x') \rightarrow ck_1(x, x') = u(x)^T u(x')$

→ $ck_1(x, x')$ is a valid kernel corresponds to feature map u

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

Similarity, $k_1(x, x')$ is a valid kernel \rightarrow exist an feature vector $\phi(x)$ such that $k_1(x, x') = \phi(x)^T \phi(x')$.

Let $v(x) = f(x)\phi(x) \rightarrow f(x)k_1(x, x')f(x') = v(x)^T v(x')$.

$\rightarrow f(x)k_1(x, x')f(x')$ is a valid kernel corresponds to feature map v

3 Problem 3: Find Kernel Matrix

$$X = \begin{bmatrix} -3 & 4 \\ 1 & 0 \end{bmatrix} \rightarrow \phi(x) = \begin{bmatrix} -3 & 4 & \sqrt{(-3)^2 + 4^2} \\ 1 & 0 & \sqrt{1^2 + 0^2} \end{bmatrix} = \begin{bmatrix} -3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix}$$

$$\begin{aligned} K &= \phi\phi^T \\ &= \begin{bmatrix} -3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} -3 & 1 \\ 4 & 0 \\ 5 & 1 \end{bmatrix} \\ &= \begin{bmatrix} (-3) \cdot (-3) + 4 \cdot 4 + 5 \cdot 5 & (-3) \cdot 1 + 4 \cdot 0 + 5 \cdot 1 \\ 1 \cdot (-3) + 0 \cdot 4 + 1 \cdot 5 & 1 \cdot 1 + 0 \cdot 0 + 1 \cdot 1 \end{bmatrix} \\ &= \begin{bmatrix} 50 & 2 \\ 2 & 2 \end{bmatrix} \end{aligned}$$