Machine Learning Lecture 8

Kernel Method

Bui Phuong Thao - 11215341 - DSEB 63

1 Exercise 1

Dual representation
Start with Ridge Regression loss function

L = $\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y) + \parallel w \parallel_2^2$

L = $\frac{1}{N}\sum_{i=1}^{N}(w^T\phi(x_n) - y_n + \parallel w \parallel_2^2$

Take derivative of L with respect to w, we have:

$\frac{\partial L}{\partial w} = \sum_{i=1}^{N}(\phi(x_n)(w^T\phi(x_n) - y_n) + \lambda w$

$\rightarrow w = \frac{-1}{\lambda}\sum_{i=1}^{N}(\phi(x_n)(w^T\phi(x_n) - y_n)$

w is a combination of inputs in featured space $(\phi(x_n) \mid 1 \leq n \leq N)$

w = $\frac{-1}{\lambda}\sum_{i=1}^{N}\phi(x_n)a_n$

w = $\frac{-1}{\lambda}\phi^T.a$ $\phi = [\phi(x_1, \phi(x_2, ...\phi(x_n]$

a = $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} w^T\phi(x_1) - y_1 \\ w^T\phi(x_2) - y_2 \\ \vdots \\ w^T\phi(x_n) - y_n \end{bmatrix}$ and $a_n = w^T\phi(x_n) - y_n$

The loss function in matrix form

$$
\begin{aligned}
L &= \frac{1}{2} \parallel \phi w - y \parallel_2^2 + \frac{\lambda}{2} \parallel w \parallel_2^2 \\
&= \frac{1}{2}(\phi w - y)^T(\phi w - y) + \frac{\lambda}{2} w^T w \\
&= \frac{1}{2}(\phi^T w^T - y^T)(\phi w - y) + \frac{\lambda}{2} w^T w \\
&= \frac{1}{2}(\phi^T w^T \phi w - w^T \phi^T y - y^T \phi w + y^T y) + \frac{\lambda}{2} w^T w \\
&= \frac{1}{2}\phi^T w^T \phi w - y^T \phi w + \frac{1}{2} y^T y + \frac{\lambda}{2} w^T w
\end{aligned}
$$

Substitute w $= \dfrac{-1}{w}\phi^T a$

$$
= \frac{1}{2\lambda^2} a^T \phi\phi^T \phi\phi^T a + \frac{1}{\lambda} y^T \phi\phi^T a + \frac{1}{2}\frac{1}{\lambda^2} a^T \phi\phi^T a
$$

$\rightarrow$ L(a,$\phi, t) = \frac{1}{2\lambda^2} a^T \phi\phi^T \phi\phi^T a - a^T \phi\phi^T t + \frac{1}{2} t^T t + \frac{\lambda}{2} a^T \phi\phi^T a$

Let K $= \phi\phi^T$ [K is symmetric]

L $= \frac{1}{2} a^T KKa - a^T Kt + \frac{\lambda}{2} a^T Ka + \frac{1}{2} t^T t$

** Proof that K is positive semi-definite:

$\Longleftrightarrow$ v$^T Kv \geq 0$ with $v \in R^n$

$v^T Kv = v^T \phi\phi^T v = \parallel v\phi^T \parallel_2^2 \geq 0$

Take the derivative of L with respect to a:

$$
\begin{aligned}
\frac{\partial L}{\partial a} = aKK - Kt + &= 0 \\
K(Ka - t + \lambda a) &= 0 \\
Ka - t + \lambda a &= 0 \\
a(K + \lambda In) &= t \\
a &= (K + \lambda In)^- 1t
\end{aligned}
$$

2 Exercise 2: Prove valid kernels

2.1 Prove that $ck_1(x, x')$ is a valid kernel

Since $k_1(x, x')$ is a valid kernel, it can be written as:

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

$$\text{k(x,x')} = ck_1(x, x') = [\sqrt{c}\phi(x)]^T[\sqrt{c}\phi(x')]$$

$\rightarrow ck_1(x, x')$ is a valid kernel

2.2 Prove that k(x, x') = f(x)k1(x, x')f(x') is a valid kernel

Since $k_1(x, x')$ is a valid kernel, it can be written as:

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

$$\text{k(x,x')} = f(x)k_1(x, x')f(x') = f(x)\phi(x)^T\phi(x')f(x')$$

$$= (f(x)\phi(x))^T(f(x')\phi(x'))$$

$$\text{With } \overline{\phi}(x) = f(x)\phi(x), k(x, x') = \overline{\phi}(x)^T\overline{\phi}(x')$$

$\rightarrow$ f(x)k1(x, x')f(x') is a valid kernel

3 Exercise 3: Compute kernel matrix

X = [(-3,4),(1,0)]
$\phi(x) = [x_1, x_2, \| x \|^T$

$$\begin{aligned}
\phi((-3,4)) &= [-3,4,5]^T \\
\phi((1,0)) &= [1,0,1]^T \\
\phi((-3,4))^T\phi((-3,4)) &= 50 \\
\phi((-3,4))^T\phi((1,0)) &= 2 \\
\phi((1,0))^T\phi((1,0)) &= 2
\end{aligned}$$

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