Kernal Method

Đạt Nguyễn Ngọc

November 2022

1 Biến đổi lại công thức từ trên lớp

$$\theta(\mathbf{x_n}) = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ ... \\ x_n \end{pmatrix} = \begin{pmatrix} \theta(x_1) \\ \theta(x_2) \\ \theta(x_3) \\ ... \\ \theta(x_n) \end{pmatrix} = \begin{pmatrix} x \\ x^2 \\ sin(x) \\ ... \\ x^3 \end{pmatrix}$$

$w^T\theta(x_0)$: predict value

Loss: $L(w) = \frac{1}{2}\sum_{n=1}^{N}(w^T\theta(x_0) - t_n)^2 + \frac{\lambda}{2}w^Tw$

Đạo hàm L(w) ta có:

$\frac{\partial L}{\partial w} = \sum_{n=1}^{N}(w^T\theta(x_0) - t_n)\theta(x_0) + \lambda w = 0$

$$\begin{aligned} w &= \frac{-1}{\lambda}\sum_{n=1}^{N}(w^T\theta(x_0) - t_n)\theta(x_0) \\ &= \sum_{n=1}^{N} a_n\theta(x_0) \\ &= \theta^T a \end{aligned}$$

in which $a_n = -\frac{1}{\lambda}(w^T\theta(x_0) - t_n)$

$$\theta = \begin{pmatrix} ...\theta(x_1)... \\ ...\theta(x_2)... \\ ...\theta(x_3)... \\ ... \\ ...\theta(x_n)... \end{pmatrix} \text{ shape of } \theta \text{ is n x d so shape of } \theta^T \text{ is d x n}$$

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ ... \\ a_n \end{pmatrix} = \begin{pmatrix} -\frac{1}{\lambda}(w^T\theta(x_1) - t_1) \\ -\frac{1}{\lambda}(w^T\theta(x_2) - t_2) \\ -\frac{1}{\lambda}(w^T\theta(x_3) - t_3) \\ ... \\ -\frac{1}{\lambda}(w^T\theta(x_n) - t_n) \end{pmatrix} \text{ shap of a is n x 1}$$

Loss:

$$\begin{aligned} L(a) &= \frac{1}{2}\sum_{n=1}^{N}(\theta a^T\theta(x_n) - t_n)^2 + \frac{\lambda}{2}\theta^T a\theta a^T \\ &= \frac{1}{2}a^T\theta\theta^T\theta\theta^T a - a^T\theta\theta t + \frac{1}{2}t^T t + \frac{\lambda}{2}a^T\theta\theta^T a \end{aligned}$$

Define Gram matrix $K = \theta\theta^T$ (Linear kernel)

$\mathbf{K} = \theta\theta^{\mathbf{T}} = \begin{pmatrix} \theta(x_1)\theta^T(x_1) & \theta(x_1)\theta^T(x_2) & ... & \theta(x_1)\theta^T(x_n) \\ ... \\ \theta(x_n)\theta^T(x_1) & \theta(x_n)\theta^T(x_2) & ... & \theta(x_n)\theta^T(x_n) \end{pmatrix}$ shape of K is n x n

$L(a) = \frac{1}{2}aKKa - a^TKt + \frac{1}{2}t^Tt + \frac{\lambda}{2}a^TKa$

Đạo hàm $L(a)$ ta có:

$\frac{\partial L}{\partial a} = KKa - Kt + \lambda Ka = 0$

$$\begin{aligned} a &= Kt(KK + \lambda K)^{-1} \\ &= Kt(K(K + \lambda I))^{-1} \\ &= KtK^{-1}(K + \lambda I)^{-1} \\ &= (K + \lambda I)^{-1} \end{aligned}$$

If we subsitute this back into the linear regression model, we obtain the following prediction for a new input x:

$$y(x) = w^T\theta(x) = a^T\theta\theta(x) = k(x)^T(K + \lambda I)^{-1}t$$