Nguyen Vu Anh Ngoc
Machine Learning 1
October 28, 2023

Homework Week 8: Kernels

Question 1. Convert primal to dual representation.

The primal problem of Ridge regression is given by:

$$\begin{aligned} J(w) &= \frac{1}{2}\sum\left(w^T\varphi\left(x_n\right)-t_n\right)^2+\frac{\lambda}{2}w^Tw \\ &= \frac{1}{2}(\varphi w-t)^T(\varphi w-t)+\frac{\lambda}{2}w^Tw \\ &= \frac{1}{2}\left(w^T\varphi^T-t^T\right)(\varphi w-t)+\frac{\lambda}{2}w^Tw \end{aligned}$$

The weight vector w is given by:

$$w=\varphi^Ta \implies w^T=a^T\varphi$$

The dual objective will have the form:

$$\begin{aligned} J(a) &= \frac{1}{2}\left(a^T\varphi\varphi^T-t^T\right)\left(\varphi\varphi^Ta-t\right)+\frac{\lambda}{2}a^T\varphi\varphi^Ta \\ &= \frac{1}{2}\left(a^T\varphi\varphi^T\varphi\varphi^Ta-a^T\varphi\varphi^Tt-t^T\varphi\varphi^Ta+t^Tt\right)+\frac{\lambda}{2}a^T\varphi\varphi^Ta \\ &= \frac{1}{2}a^T\varphi\varphi^T\varphi\varphi^Ta-a^T\varphi\varphi^Tt+\frac{1}{2}t^Tt+\frac{\lambda}{2}a^T\varphi\varphi^Ta \\ &= \frac{1}{2}a^TKKa-a^TKt+\frac{1}{2}t^Tt+\frac{\lambda}{2}a^TKa \end{aligned}$$

Question 2. Prove valid kernels.

A **valid** kernel corresponds to a scalar product in some (perhaps infinite dimensional) feature space.

$$k\left(x,x'\right)=\langle\varphi^T(x),\varphi(x')\rangle$$

Equation 6.13

$$\begin{aligned} k'\left(x,x'\right) &= c\cdot k\left(x,x'\right) \\ &= c\cdot\langle\varphi(x),\varphi\left(x'\right)\rangle \\ &= \left\langle\sqrt{c}\varphi(x),\sqrt{c}\varphi\left(x'\right)\right\rangle \\ &= \langle\varphi'(x),\varphi'\left(x'\right)\rangle \text{ (q.e.d)} \end{aligned}$$

Equation 6.14

$$\begin{aligned} k'\left(x,x'\right) &= f(x)k\left(x,x'\right)f\left(x'\right) \\ &= f(x)\left\langle\varphi(x),\varphi\left(x'\right)f\left(x'\right)\right. \\ &= \langle f(x),\varphi(x)\rangle\left\langle f\left(x'\right),\varphi\left(x'\right)\right\rangle \\ &= \langle\varphi'(x),\varphi'\left(x'\right)\rangle \text{ (q.e.d)} \end{aligned}$$

Question 3. Given data set X and feature map φ. Calculate the kernel matrix K.

Given:

$$X = \begin{bmatrix} -3 & 4 \\ 1 & 0 \end{bmatrix}, \quad \varphi(x) = \begin{bmatrix} x_1 \\ x_2 \\ \|x\| \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \sqrt{x_1^2 + x_2^2} \end{bmatrix}$$

Apply the feature map to X:

$$\varphi = \begin{bmatrix} -3 & 4 & 5 \\ 1 & 0 & 1 \end{bmatrix}$$

Calculate kernel matrix K:

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