Nguyen Vu Anh Ngoc
Machine Learning 1
November 2, 2023

Homework Week 8: Kernels

Question 1. Convert primal to dual representation.

The primal problem of Ridge regression is given by:

$$\begin{aligned} J(\mathbf{w}) &= \frac{1}{2}\sum\left(\mathbf{w}^T\mathbf{\Phi}(\mathbf{x}_n) - t_n\right)^2 + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \\ &= \frac{1}{2}(\mathbf{\Phi w} - \mathbf{t})^T(\mathbf{\Phi w} - \mathbf{t}) + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \\ &= \frac{1}{2}(\mathbf{w}^T\mathbf{\Phi}^T - \mathbf{t}^T)(\mathbf{\Phi w} - \mathbf{t}) + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \end{aligned}$$

Derivation of $J(\mathbf{w})$:

$$\nabla J(\mathbf{w}) = \mathbf{\Phi}^T(\mathbf{\Phi w} - \mathbf{t}) + \lambda\mathbf{w} = 0$$
$$\implies \mathbf{w} = \mathbf{\Phi}^T\mathbf{a}$$

The dual objective will have the form:

$$\begin{aligned} J(\mathbf{a}) &= \frac{1}{2}(\mathbf{a}^T\mathbf{\Phi\Phi}^T - \mathbf{t}^T)(\mathbf{\Phi\Phi}^T\mathbf{a} - \mathbf{t}) + \frac{\lambda}{2}\mathbf{a}^T\mathbf{\Phi\Phi}^T\mathbf{a} \\ &= \frac{1}{2}\left(\mathbf{a}^T\mathbf{\Phi\Phi}^T\mathbf{\Phi\Phi}^T\mathbf{a} - \mathbf{a}^T\mathbf{\Phi\Phi}^T\mathbf{t} - \mathbf{t}^T\mathbf{\Phi\Phi}^T\mathbf{a} + \mathbf{t}^T\mathbf{t}\right) + \frac{\lambda}{2}\mathbf{a}^T\mathbf{\Phi\Phi}^T\mathbf{a} \\ &= \frac{1}{2}\mathbf{a}^T\mathbf{KKa} - \mathbf{a}^T\mathbf{Kt} + \frac{1}{2}\mathbf{t}^T\mathbf{t} + \frac{\lambda}{2}\mathbf{a}^T\mathbf{Ka} \end{aligned}$$

Differentiate $J(\mathbf{a})$ with respect to $\mathbf{a}$:

$$\frac{\partial J(\mathbf{a})}{\partial \mathbf{a}} = \mathbf{KKa} - \mathbf{Kt} + \lambda\mathbf{Ka}$$

Setting the gradient to zero, we have:

$$(\mathbf{KK} + \lambda\mathbf{K})\mathbf{a} = \mathbf{Kt}$$

Finally, we have solution for $\mathbf{a}$:

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

Question 2. Prove valid kernels.

A **valid** kernel corresponds to a scalar product in some (perhaps infinite dimensional) feature space. *(PRML, p.294)*

$$k\left(x, x'\right) = \langle\Phi(x), \Phi(x')\rangle$$

Equation 6.13

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

Equation 6.14

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

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

Given:

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

Apply the feature map to X:

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

Calculate kernel matrix K:

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