Kernel Methods

Setup: Given $\mathcal{D} := \{x_i, y_i\}$, $x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$, $i = 1, \dots, n$

Linear regression:
$$y_i = f(x_i) = \sum_{i=1}^{m} \theta_i \varphi_i(x_i) = \langle \theta, \varphi(x_i) \rangle$$

$$\varphi : \mathbb{R}^d \longrightarrow \mathbb{R}^m, \quad \varphi(x) = (\varphi_1(x), \dots, \varphi_m(x))$$

Training:
$$\theta^* = \underset{\theta \in \Theta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \ell(y_i, \langle \theta, \varphi(x_i) \rangle) + \frac{\lambda}{2} \|\theta\|^2 \quad (1)$$

MAP estimate: $$\theta^* = \underbrace{(\Phi^T \Phi - \lambda I)}_{m \times m}^{-1} \Phi^T y$$

$$\underset{n \times m}{\Phi} = \begin{bmatrix} \varphi_1(x_1) & \cdots & \varphi_m(x_1) \\ \vdots & & \\ \varphi_1(x_n) & \cdots & \varphi_m(x_n) \end{bmatrix}$$

Prediction:
$$y^* = \langle \theta^*, \varphi(x^*) \rangle = \varphi(x^*) \underbrace{(\Phi^T \Phi - \lambda I)^{-1} \Phi^T y}_{\theta^*}$$

Representers theorem:

The minimum of (1) can be obtained if θ takes the following form:
$$\underset{\in \mathbb{R}^m}{\theta} = \sum_{i=1}^{n} \alpha_i \varphi(x_i), \quad \alpha \in \mathbb{R}^n$$

n ... $\lambda \|\theta\|^2$

$$\inf_{\vartheta} \frac{1}{n} \sum_{i=1} \ell(y_i, \langle \vartheta, \varphi(x_i) \rangle) + \frac{\lambda}{2} \|\vartheta\|$$

equivalent $\Longrightarrow$

$$\left\{ \begin{array}{l} \inf\limits_{\alpha \in \mathbb{R}^n} \frac{1}{n} \sum_{i=1}^{n} \ell(y_i, \langle \vartheta, \varphi(x_i) \rangle) + \frac{\lambda}{2} \|\vartheta\|^2 \\ \text{s.t.} \quad \boxed{\vartheta = \sum_{i=1}^{n} \alpha_i \varphi(x_i)} \end{array} \right\} \quad (2)$$

<u>Kernel function</u>: $\boxed{K(x, x') = \langle \varphi(x), \varphi(x') \rangle}$

- symmetric
- PSD

Finite dim: $\underline{K_{ij}} = \sum_{i=1}^{m} \varphi_m(x_i) \varphi_m(x_j)$ $\xrightarrow{m \to +\infty}$ $\boxed{K(x, x') = \int_0^{\infty} \varphi_m^T(x) \varphi_m}$ inf. dim

$m \times m$

$$\boxed{\bullet \; \langle \vartheta, \varphi(x_j) \rangle = \boxed{\sum_{i=1}^{n} \alpha_i K(x_i, x_j)} = (K\alpha)_j}$$

$$\bullet \; \|\vartheta\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \underbrace{\langle \varphi(x_i), \varphi(x_j) \rangle}_{K_{ij}} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j K_{ij} = \alpha^T K \alpha$$

$$(2) \Longrightarrow \boxed{\inf_{\underline{\alpha \in \mathbb{R}^n}} \frac{1}{n} \sum_{i=1}^{n} \ell(y_i, (K\alpha)_i) + \frac{\lambda}{2} \alpha^T K \alpha} \quad (3)$$

e.g. $K(x, x') = \exp\left(-\frac{1}{2} \frac{(x - x')^2}{\ell^2}\right)$, RBF

$K(x, x') = \left(\sum_{i=1}^{d} x_i x_i'^T \right)$, polynomial kernel

Recall, Bayesian linear regression: $\boxed{y = \underline{f(x)} + \varepsilon}$, $\cancel{\langle \theta, \phi(x) \rangle}$

- $p(y | X, \theta) = \mathcal{N}(y | \langle \underline{\theta}, \phi(x) \rangle, \sigma^2 I_{n \times n})$, likelihood

- $p(\theta) = \mathcal{N}(\theta | 0, I)$, prior

$m \to \infty$, $\theta \in \Theta$ inf. dim Hilbert space.

$$p(\theta | X, y) = \frac{p(y | X, \theta)\, p(\theta)}{\int p(y | X, \theta)\, p(\theta)\, d\theta}$$

$$\xrightarrow{m \to +\infty} \boxed{p(f | X, y) = \frac{p(y | f, X)\, p(f)}{\int p(y | X, f)\, p(f)\, df}} \checkmark$$