Kernel Methods

1. Domain $X = D \subseteq \mathbb{R}^p$, D closed & bounded.

2. Goal: find $f(x) : D \longmapsto \mathbb{R}$ s.t. $f(x_i) \simeq y_i$

3. Candidates for f, $f \in (\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$

   (a) If $\mathcal{H}$ is reproducing KHS. $\exists K$ s.t.

$$f(x) = \langle f(\cdot), K(\cdot, x) \rangle_{\mathcal{H}}$$

Ex: (Cubic Splines) $\mathcal{H} = W^{(2)} =$ square int. twice diverentiable

Ex: $D = \mathbb{R}^p$ $K(x, y) = x^T y + 1$

Corresponding RKHS $\mathcal{H} = (\beta^T x + \beta_0 \mid \beta \in \mathbb{R}^p, \beta_0 \in \mathbb{R}\}$

Ex: Polynomial Kernel $K(x, y) = (x^T + y^T)^d$ $d = 1, 2 ..$

$\mathcal{H} = \{$ all polynomials $p(x)$ with $\deg(p) \leq d$

Observation $K(x, y) = \sum_{i=1}^{m} h_m(x) h_m(y)$ psd kernel function

Define $h(x) = \begin{pmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{pmatrix}$ and consider the feature map

$$T = \{(x_i, y_i)\}_{i=1}^{n} \longmapsto T^* \{ h(x_i), y_i \}_{i=1}^{N}$$

<u>Representer Thm:</u> Tells us $f(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i)$

So $\hat{f} = \underset{f \in H_K}{\arg\min} \sum_{i=1}^{n} L(y_i, f(x_i)) + \lambda \|f\|_{H_K}^2$

$\Rightarrow \hat{\alpha} = \underset{\alpha}{\arg\inf} \sum_{i=1}^{n} L(y_i, y_i (\underset{\sim}{K}\alpha)_i) + \lambda \alpha^T K \alpha$

Under the new training set $T^* = \{\overbrace{h(x_i)}^{z_i}, y_i\}_{i=1}^{n}$

Using $K(x,y) = x^T y$ $\mathcal{H}_K = \{ f(x) = \beta^T x \}$

now: $\hat{f}^* = \underset{f^* \in H_K^*}{\arg\inf} \sum_{i=1}^{N} L(y_i, f^*(h(x_i))) + \lambda \|f^*\|_{H_K^*}$

$\hat{f}^*(x) = \sum_{i=1}^{n} \alpha_i^* K^*(x, x_i) \Rightarrow \alpha_i^* = \alpha_i$

<u>Moral:</u> Using $K(x,y) = \sum_{m=1}^{M} h_m(x) h_m(y)$

is equivalent to using the feature map $h: \mathbb{R}^p \mapsto \mathbb{R}^m$

$x_i \mapsto h(x_i)$