Representer theorem for RKHS

$K: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a PDS kernel

H is its RKHS ← Moore - Aron... theorm

$g: \mathbb{R} \to \mathbb{R}$ increasing function

$l: \mathbb{R}^m \to \mathbb{R}$ loss function

$$\underset{h \in H}{\operatorname{argmin}} \; g(\|h\|_H) + \underbrace{l(h(x_1), \dots, h(x_m))}$$

$$\in \operatorname{span}(\{K(x_i, \cdot) : i \in [m]\})$$

Proof: $H_S = \{ \sum_{i=1}^{m} a_i K(x_i, \cdot) : a_i \in \mathbb{R}, x_i \in S \}$

Any $h \in H$ can be written as

$h = h_0 + h_0^\perp$ where $h_0 \in H_S$ and $h_0^\perp \in H_S^\perp$

By reproducing property,

$$\begin{aligned} h(x_i) &= \langle h, k(x_i, \cdot) \rangle \\ &= \langle h_0 + h_0^\perp, k(x_i, \cdot) \rangle \\ &= \langle h_0, k(x_i, \cdot) \rangle + \underset{0}{\underset{\|}{\langle h_0^\perp, k(x_i, \cdot) \rangle}} \end{aligned}$$

($h_0^\perp$ is ortho to any function in H_S)

$$\begin{aligned} &= \langle h_0, k(x_i, \cdot) \rangle \\ &= h_0(x_i) \end{aligned}$$

$$l(h(x_1), \dots, h(x_m)) = l(h_0(x_1), \dots, h_0(x_m))$$

$$\|h\| = \|h_0 + h_0^\perp\|$$

$$\left(\|h_0 + h_0^\perp\|^2 = \|h_0\|^2 + \|h_0^\perp\|^2 + \underbrace{2\langle h_0, h_0^\perp \rangle}_{} \right)$$

$$= \|h_0\|^2 + \|h_0^\perp\|^2$$

if $\|h\| > \|h_0\|$, then

$g(\|h\|) > g(\|h_0\|)$ (g is increasing)

$$\Rightarrow \quad \underset{h \in H}{\operatorname{argmin}} \; l(h(x_1), \dots, h(x_m)) + g(\|h\|)$$

$$= \underset{h \in H_S}{\operatorname{argmin}} \; l(h(x_1) \dots, h(x_m)) + g(\|h\|)$$

FoML Mohri 6.22

Anomaly detection applications

→ given a data point, is it an outlier?
e.g. climate, credit card fraud
without knowing $\mathcal{D}$.

→ Comparing data sets

a) $$\mathcal{L}(r, c, \alpha) = r^2 + \sum_{i=1}^{m} \alpha_i \left(\|\Phi(x_i) - c\|^2 - r^2 \right)$$

<u>KKT</u>

1) $\partial_r \mathcal{L}(r, c, \alpha) = 0$ ✓

2) $\partial_c \mathcal{L}(r, c, \alpha) = 0$ ←

3) Complementarity

$$\sum_{i=1}^{m} \alpha_i \left(\underbrace{\|\Phi(x_i) - c\|^2 - r^2}_{} \right) = 0$$

$$\alpha_i \geq 0$$

For each i,

$$\alpha_i \left(\|\Phi(x_i) - c\|^2 - r^2 \right) = 0$$

From 1), $\sum_{i=1}^{m} \alpha_i = 1$

$$H_S = \{ \sum_{i=1}^{m} a_i \Phi(x_i) : a_i \in \mathbb{R}, x_i \in S \}$$

For any $c \in H$, $c = c_0 + c_0^\perp$

$$\begin{aligned} &\|\Phi(x_i) - c_0 - c_0^\perp\|^2 \\ &= \|\Phi(x_i) - c_0\|^2 + \|c_0^\perp\|^2 + 2\langle \underset{H_S}{\underset{\uparrow}{\Phi(x_i)}} - \underset{H_S}{\underset{\cap}{c_0}}, c_0^\perp \rangle \\ &= \|\Phi(x_i) - c_0\|^2 + \cancel{\|c_0^\perp\|^2} \end{aligned}$$

$$\underset{r, a \in \mathbb{R}^m}{\operatorname{argmin}} \; r^2 \text{ s.t. } \left(\|\Phi(x_i) - \sum_{j=1}^{m} a_j \Phi(x_j)\|^2 - r^2 \right) \leq 0$$

$a = [a_1, \dots, a_m]^T$