· Functional : function $\mapsto \mathbb{R}$

eg.

evaluation functional : $E_x(\cdot)$ is evaluating at x $E_x(f) = f(x)$

summation functional : $S_{x_1,\dots,x_n}(f) = \sum_i f(x_i)$

integration functional : $I_{[a,b]}(f) = \int_a^b f(x)\,dx$

the compositional of a function $g: \mathbb{R} \mapsto \mathbb{R}$ with a functional is also a functional

· Derivative of functionals

let k be a kernel, $\mathcal{H}$ be the RKHS of k, $E: \mathcal{H} \mapsto \mathbb{R}$ be a functional

$DE(f): \mathcal{H} \to \mathcal{H}$ is the functional gradient of E at f

s.t. $$\lim_{\|h\|_{\mathcal{H}} \to 0} \frac{\|E(f+h) - E(f) - \langle DE(f), h\rangle_{\mathcal{H}}\|}{\|h\|_{\mathcal{H}}} = 0 \quad f \in \mathcal{H} \quad g \in \mathcal{H}$$

eg. $E_x(f) = f(x)$

$E_x(f+h) = f(x) + h(x) = f(x) + \langle h, k(\cdot, x)\rangle_{\mathcal{H}}$

$DE_x(f) = k(x, \cdot)$

$$\lim_{\|h\|_{\mathcal{H}} \to 0} \frac{\| E_x(f+h) - E_x(f) - \langle DE(f), h\rangle_{\mathcal{H}} \|}{\|h\|_{\mathcal{H}}} = \lim_{\|h\|_{\mathcal{H}} \to 0} \frac{| f(x) + h(x) - f(x) - \langle k(x,\cdot), h\rangle |}{\|h\|_{\mathcal{H}}} = 0$$

eg. $E(f) = \|f\|_{\mathcal{H}}^2$

$E(f+h) = \|f+h\|_{\mathcal{H}}^2 = \|f\|_{\mathcal{H}}^2 + \|h\|_{\mathcal{H}}^2 + 2\langle f, h\rangle_{\mathcal{H}}$

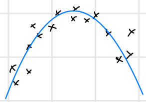
$DE(f) = 2f$

$$\lim_{\|h\|_{\mathcal{H}} \to 0} \frac{\| E(f+h) - E(f) - \langle DE(f), h\rangle \|}{\|h\|_{\mathcal{H}}} = \lim_{\|h\|_{\mathcal{H}} \to 0} \frac{\|h\|_{\mathcal{H}}^2}{\|h\|_{\mathcal{H}}} = 0$$

let $E: \mathcal{H} \mapsto \mathbb{R}$ $g: \mathbb{R} \to \mathbb{R}$ $h(f) = g(E(f))$

$Dh(f) = g'(E(f)) \cdot D_E(f)$ chain rule

Eg. Kernel Regression



given data $\{(x_i, y_i)\}_{i=1}^{m}$

model: $f = \sum_{i=1}^{m} \alpha_i k(x_i, \cdot) \in \mathcal{H}$

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|_2^2}{2}\right)$$

$$\min_{f \in \mathcal{H}} \mathcal{L}(f) = \frac{1}{m} \sum_{i=1}^{m} \left(f(x_i) - y_i\right)^2 + \lambda \|f\|_{\mathcal{H}}^2$$

$$D\mathcal{L}(f) = \frac{2}{m} \sum_{i=1}^{m} \left(f(x_i) - y_i\right) \cdot k(x_i, \cdot) + 2\lambda f$$

$$= \frac{2}{m} \sum_{i=1}^{m} \left[\left(f(x_i) - y_i\right) + 2\lambda \alpha_i\right] k(x_i, \cdot) \qquad \frac{d\mathcal{L}}{d\alpha_i} = \frac{2}{m}\left(f(x_i) - y_i\right) + 2\lambda\alpha_i$$

$$\|D\mathcal{L}(f)\|_{\mathcal{H}}^2 = \frac{4}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left(f(x_i) - y_i\right) k(x_i, x_j) \left(f(x_j) - y_j\right)$$

$$+ 4\lambda^2 \|f\|_{\mathcal{H}}^2$$

$$+ 2 \left\langle \frac{2}{m} \sum_{i=1}^{m} \left(f(x_i) - y_i\right) k(x_i, \cdot), 2\lambda f \right\rangle_{\mathcal{H}}$$

$$= \frac{4}{m^2} \left(f(x) - y\right)^T K \left(f(x) - y\right) + 4\lambda^2 \alpha^T K \alpha + \frac{8}{m} \left(f(x) - y\right)^T f(x)$$

Backtracking line search

while $\mathcal{L}(f) - \gamma t \|D\mathcal{L}(f)\|_{\mathcal{H}}^2 < \mathcal{L}(f - tD\mathcal{L}(f))$

$t := 0.5t$