Statistical Techniques in Robotics (16-831, F08) Lecture #22 (Nov. 6)

Reproducing Kernel Hilbert Space

Lecturer: Drew Bagnell *Scribe: Bryan Low*

The Reproducing Kernel Hilbert Space (RKHS), denoted by $\mathcal{H}_k$, is the space of functions that can be written as $\sum_i \alpha_i k(x_i, x)$. For RKHS, linearity holds, that is, if $f \in \mathcal{H}_k$ and $g \in \mathcal{H}_k$, $\alpha f + \beta g \in \mathcal{H}_k$.

The *inner product* of $f, g \in \mathcal{H}_k$ is defined as

$$\langle f, g \rangle \triangleq \sum_i \sum_j \alpha_i \beta_j k(x_i, x_j) = \boldsymbol{\alpha}^\top \mathbf{K} \boldsymbol{\beta}$$

where $f(.) = \sum_i \alpha_i k(x_i, .)$, $g(.) = \sum_j \beta_j k(x_j, .)$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors comprising, respectively, α_i and β_i components, and $\mathbf{K}$ is a matrix comprising $k(x_i, x_j)$ components. Further, $||f||^2 \triangleq \langle f, f \rangle = \boldsymbol{\alpha}^\top \mathbf{K} \boldsymbol{\alpha}$. Note that $\mathbf{K}$ must satisfy the positive definite (hence, norm is positive) and symmetric (i.e., $k(x_i, x_j) = k(x_j, x_i)$) conditions.

For example, if $f(.) = k(x_i, .)$ and $g(.) = k(x_j, .)$, $\langle f, g \rangle = k(x_i, x_j)$. Assuming that the kernels are Gaussian, if they are far apart, the inner product tends to 0. If they are very close, the inner product tends to $k(x_i, x_i)$ (often assumed to be 1); the inner product becomes -1 if $g(.) = -k(x_j, .)$.

The loss functional $L[f]$ is defined as

$$L[f] \triangleq \sum_i (f(x_i) - y_i)^2 + \frac{\lambda}{2} ||f||_k^2 \ .$$

If the update rule is of the form $f \leftarrow f - \alpha \nabla L$, we need to find the expression for ∇L.

A functional $F : f \rightarrow \mathbb{R}$ accepts a function f as input and returns a real number as output. Using local linearization, we have

$$F[f + \varepsilon g] = F[f] + \varepsilon \langle \nabla F, g \rangle + \mathcal{O}(\varepsilon^2) \ .$$

For example, if $F[f] = ||f||^2$, $\nabla F = 2f$.

Proof.

$$\begin{aligned} F[f + \varepsilon g] &= \langle f + \varepsilon g, f + \varepsilon g \rangle \\ &= \langle f, f \rangle + 2 \langle f, \varepsilon g \rangle + \varepsilon^2 \langle g, g \rangle \\ &= ||f||^2 + \varepsilon \langle 2f, g \rangle + \varepsilon^2 ||g||^2 \end{aligned} \qquad \square$$

Supposing $L[f] = ||f||^2$ (i.e., $\nabla L = 2f$), the update rule becomes

$$f \leftarrow f - 2\alpha f = (1 - 2\alpha) f \ ,$$

which shrinks the function f.

The reproducing property of RKHS can be observed from

$$\begin{aligned} F_{x_i}[f] &\triangleq f(x_i) \\ &= \textstyle\sum_j \alpha_j k(x_j, x_i) \\ &= \langle f, k(x_i, .) \rangle \ . \end{aligned}$$