1 Warm-up

Let $\mathcal{H}$ be a RKHS associated with a positive definite kernel $k : \mathsf{X} \times \mathsf{X} \to \mathbb{R}$.

1. Prove that for all $(x, y) \in \mathsf{X} \times \mathsf{X}$ and $f \in \mathcal{H}$,

$$|f(x) - f(y)| \leqslant \|f\|_{\mathcal{H}} \|k(x, \cdot) - k(y, \cdot)\|_{\mathcal{H}} .$$

2. Prove that the kernel k associated with $\mathcal{H}$ is unique, i.e. if $\widetilde{k}$ is another potitive definite kernel satisfying the RKHS properties for $\mathcal{H}$, then $k = \widetilde{k}$.

3. Prove that for all $x \in \mathsf{X}$, the function defined on $\mathcal{H}$ by $\delta_x : f \mapsto f(x)$ is continuous.

2 Kernel Ridge regression

Let $\mathcal{H}$ be a RKHS on $\mathcal{X}$ with kernel k. We consider the regression model $Y_i = f^*(X_i) + \xi_i$, $i \in \{1, \ldots, n\}$, with ξ_i, $1 \leq i \leq n$, independent centered noise with finite variance. The unknown function f^* is estimated by the solution $\widehat{f}$ of the convex minimization problem

$$\widehat{f} = \operatorname*{argmin}_{f \in \mathcal{H}} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(X_i))^2 + \frac{\lambda}{n} \|f\|_{\mathcal{H}}^2 \right\} ,$$

with $\lambda > 0$.

2.1 Solving Kernel ridge regression

1. Check that $\widehat{f} : x \mapsto \sum_{j=1}^{n} \widehat{\beta}_j k(X_j, x)$ where $\widehat{\beta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_n)^\top$ is solution to

$$\widehat{\beta} = \operatorname*{argmin}_{\beta \in \mathbb{R}^n} \left\{ \|Y - K\beta\|^2 + \lambda \beta^\top K \beta \right\}$$

   with K defined by $K = (k(X_i, X_j))_{1 \leq i,j \leq n}$. Comment on this result.

2. Assume that K is non-singular. Give an explicit expression for $\widehat{\beta}$.

2.2 Bias and variance

We assume that $f^* \in \mathcal{H}$ and we write

$$f_V^* : x \mapsto \sum_{i=1}^{n} \beta_i^* k(X_i, x)$$

for the projection of f^* onto the linear span $V = \mathrm{span}\{k(X_i, .) : i = 1, \ldots, n\}$, with respect to the Hilbert norm $\| \cdot \|_{\mathcal{H}}$. We write $K = \sum_{i=1}^{n} \lambda_i u_i u_i^\top$ for an eigenvalue decomposition of K.

1. Check that

$$K\widehat{\beta} = \sum_{i=1}^{n} \frac{\lambda_i}{\lambda_i + \lambda} \langle Y, u_i \rangle u_i \quad \text{with} \quad Y = (Y_1, \ldots, Y_n)^\top .$$

2. Check that
$$\|\mathbb{E}[K\widehat{\beta}] - K\beta^*\|_2^2 = \sum_{i=1}^{n} \left(\frac{\lambda\lambda_i}{\lambda_i + \lambda} \right)^2 \langle \beta^*, u_i \rangle^2.$$

3. We assume henceforth that the $\xi_i = Y_i - f^*(X_i)$, $i = 1, \ldots, n$, have a covariance $\mathbb{V}[\xi] = \sigma^2 I_n$. Check that the covariance matrix of $K\widehat{\beta}$ is equal to
$$\mathbb{V}[K\widehat{\beta}] = \sum_{i=1}^{n} \left(\frac{\lambda_i \sigma}{\lambda_i + \lambda} \right)^2 u_i u_i^\top .$$

4. We define $\|f\|_n^2 := \frac{1}{n}\sum_{i=1}^{n} f(X_i)^2$. Prove that
$$\mathbb{E}\left[\|\widehat{f} - f^*\|_n^2 \right] = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\lambda_i}{\lambda + \lambda_i} \right)^2 \left(\lambda^2 \langle \beta^*, u_i \rangle^2 + \sigma^2 \right).$$