Theory of Machine Learning

Exercise sheet 13 — Session 13

Exercise I (Woodbury matrix identities) ✎. Given the real-valued block matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}$$

with blocks $A \in \mathbb{R}^{m\times m}$, $D \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times n}$, and $C \in \mathbb{R}^{n\times m}$. The objective of this exercise is to derive the Woodbury matrix identities, used in the lecture slide 162, by applying Gaussian elimination in block form to M.

Let us make the simplifying assumption that A and D are invertible. Also, we introduce the Schur complement M/A of block A defined as $M/A := D - CA^{-1}B$ (assume that M/A is invertible). In the same, fashion we define $M/D := A - BD^{-1}C$ (also, considered invertible).

1. Compute, by Gaussian elimination, the inverse M^{-1} where you start by setting the block C to zero. One should obtain:

$$M^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(M/A)^{-1}CA^{-1} & -A^{-1}B(M/A)^{-1} \\ -(M/A)^{-1}CA^{-1} & (M/A)^{-1} \end{pmatrix}.$$

2. Repeat the Gaussian elimination, but this time begin by first setting the upper-right block B to zero and obtain

$$M^{-1} = \begin{pmatrix} (M/D)^{-1} & -(M/D)^{-1}BD^{-1} \\ -D^{-1}C(M/D)^{-1} & D^{-1} + D^{-1}C(M/D)^{-1}BD^{-1} \end{pmatrix}.$$

3. Show the following identities

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1},$$
$$(D - CA^{-1}B)^{-1} = D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}.$$

4. Show that

$$(A - BD^{-1}C)^{-1}B = A^{-1}B(D - CA^{-1}B)^{-1}D.$$

5. Show that

$$(I + BB^\top)^{-1} = I - B(I + B^\top B)^{-1}B^\top,$$
$$(I + BB^\top)^{-1}B = B(I + B^\top B)^{-1}.$$

Exercise II (Representer theorem) ✎. Let us recall the Representer theorem: let $\mathcal{H}$ be the RKHS associated to k defined on $\mathcal{X}$. Let $S = \{x_1, ..., x_n\} \subset \mathcal{X}$ be a finite set of points. Let $\Psi : \mathbb{R}^{n+1} \to \mathbb{R}$ be a function, strictly increasing in the last variable. Then any solution to the minimization problem

$$\arg\min_{f\in\mathcal{H}} \Psi(f(x_1), \dots, f(x_n), \|f\|_{\mathcal{H}})$$

admits a representation of the form

$$\exists(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n, \forall x \in \mathcal{X}, \qquad f(x) = \sum_{i=1}^n \alpha_i k(x_i, x).$$

The objective is to prove the theorem, meaning

$$\arg\min_{f\in\mathcal{H}} \Psi(f(x_1), \dots, f(x_n), \|f\|_{\mathcal{H}}) = \arg\min_{f\in\operatorname{span}\{K_{x_i} : i\in[\![n]\!]\}} \Psi(f(x_1), \dots, f(x_n), \|f\|_{\mathcal{H}}).$$

1. As $\mathcal{S} := \text{span}\{K_{x_i} : i \in [\![n]\!]\}$ is a finite-dimensional subspace, therefore any function $f \in \mathcal{H}$ can be uniquely decomposed as
$$f = f_\mathcal{S} + f_\perp \,,$$
with $f_\mathcal{S} \in \mathcal{S}$ and $f_\perp \perp \mathcal{S}$.
Show that
$$\forall i \in [\![n]\!], \qquad f(x_i) = f_\mathcal{S}(x_i) \,.$$
2. Given the Pythagoras' theorem in $\mathcal{H}$ $\left(\|f\|_\mathcal{H}^2 = \|f_\mathcal{S}\|_\mathcal{H}^2 + \|f_\perp\|_\mathcal{H}^2\right)$, show that
$$\forall f \in \mathcal{H}, \qquad \Psi(f(x_1), \dots, f(x_n), \|f\|_\mathcal{H}) \geq \Psi(f_\mathcal{S}(x_1), \dots, f_\mathcal{S}(x_n), \|f_\mathcal{S}\|_\mathcal{H}) \,,$$
with equality if and only if $\|f_\perp\|_\mathcal{H} = 0$.
3. Given the previous question, is it possible that a minimizer f of Ψ is not in the span $\mathcal{S}$?