"Kernel methods in machine learning"
Homework 2

1 General instructions:

1. **The delivery must be a single PDF file containing answers to all questions.**

2. **You must upload the PDF to the GradeScope platform after creating an account there. See instructions in the course webpage for using GradeScope.**

This homework contains both mathematical and coding questions:

- Some questions require providing equations **without proofs** (see exercise 3). In that case, please only include the final results, without the details of the derivations. For exercises 1 and 2 you must provide proofs.

- Exercises 3 essentially coding questions. These require implementing some methods that are described in a Jupyter notebook (Homework.ipnb) attached to this homework. Please follow the template of the notebook and only fill in the gaps whenever asked for. You should **take a screenshot of the code** you wrote and include it to the PDF.

- Some questions require running a code block in the Jupyter notebook to check your implementation. You should take a screenshot of all whole output of the (figures + any text that appears) and include it to the PDF.

Exercice 1. Sobolev spaces

Let $\mathcal{H} = \{f : [0,1] \to \mathbb{R}\,,\ \text{absolutely continuous}, f' \in L^2([0,1]), f(0) = 0\}$, endowed with the bilinear form

$$\forall f, g \in \mathcal{H}\,, \quad \langle f, g\rangle_{\mathcal{H}} = \int_0^1 (f(u)g(u) + f'(u)g'(u))\, du\,.$$

Show that $\mathcal{H}$ is an RKHS, and compute its reproducing kernel.

Exercice 2. Gaussian RKHS

For any $\sigma > 0$, let K_σ be the normalized Gaussian kernel on $\mathbb{R}^d$:

$$\forall x, y \in \mathbb{R}^d \quad K_\sigma(x,y) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^d} \exp\left(-\frac{\| x - y \|^2}{2\sigma^2}\right)\,,$$

and let $\mathcal{H}_\sigma$ be its reproducing kernel Hilbert space (RKHS).

1. Recall a proof of the positive definiteness of K.
2. For any $0 < \sigma < \tau$, show that
$$\mathcal{H}_\tau \subset \mathcal{H}_\sigma \subset L_2(\mathbb{R}^d)\,,$$
3. For any $0 < \sigma < \tau$ and $f \in \mathcal{H}_\tau$, show that
$$\| f \|_{\mathcal{H}_\tau} \geq \| f \|_{\mathcal{H}_\sigma} \geq \| f \|_{L_2(\mathbb{R}^d)}\,,$$
and that
$$0 \leq \| f \|^2_{\mathcal{H}_\sigma} - \| f \|^2_{L_2(\mathbb{R}^d)} \leq \frac{\sigma^2}{\tau^2}\left(\| f \|^2_{\mathcal{H}_\tau} - \| f \|^2_{L_2(\mathbb{R}^d)}\right)\,.$$
4. For any $\tau > 0$ and $f \in \mathcal{H}_\tau$, show that
$$\lim_{\sigma \to 0} \| f \|_{\mathcal{H}_\sigma} = \| f \|_{L_2(\mathbb{R}^d)}\,.$$

Exercice 3. Support Vector Classifier

Consider a dataset of N pairs (x_i, y_i) where each x_i is a vector of dimension d

and y_i is a binary class, i.e. $y_i \in \{-1, 1\}$. We would like to separate the two classes of samples with a **separating hyper-surface** of equation $f(x_i)+b = 0$ such that $f(x_i)+b \leq 0$ if x_i belongs to the class $y_i = -1$ and $f(x_i)+b \geq 0$ if $y_i = 1$. To achieve this, we consider functions f that belong to a Reproducing Kernel Hilbert Space $\mathcal{H}$ of kernel k. Such choice allows to represent highly non-linear hyper-surfaces while still solving a convex problem of the form:

$$\begin{aligned} \min_{f,b,\xi_i} \quad & \frac{1}{2}\|f\|^2 + C\sum_{i=1}^{n} \xi_i \\ \text{s.t.} \quad & y_i(f(x_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{aligned} \tag{1}$$

1. **Without proofs**:
   (a) Provide an expression for the Lagrangian of the problems in eq. (1) in terms of N dual parameters $\alpha_i \geq 0$ corresponding the margin inequalities and N dual parameters $\mu_i \geq 0$ corresponding to the positivity constraints on ξ_i whenever applicable.
   (b) Using the optimality condition on the Lagrangian, express the dual problem as a **constrained minimization** over $(\alpha_i)_{i\in\{1,\dots,N\}}$ and express $f(x)$ in terms of α_i and relevant quantities.
   (c) Using Strong duality (KKT conditions), find a condition characterizing the **support vector** points x_i that are on the margin of the separating hyper-surface, i.e. the points satisfying the equation $y_i(f(x_i) + b) = 1$.

2. **(a)** In the notebook, implement the method `kernel` of the classes RBF and Linear, which takes as input two data matrices X and Y of size $N \times d$ and $M \times d$ and returns a gramm matrix G of shape $N \times M$ whose components are $k(x_i, y_j) = \exp(-\|x_i - y_i\|^2/(2\sigma^2))$ for RBF and $k(x_i, y_j) = x_i^\top y_j$ for the linear kernel. (The fastest solution does not use any for loop!)
   In the notebook, the class `KernelSVC` corresponds to eq. (1):
   (b) Implement the method `fit` that computes the optimal dual parameters α_i, the parameter b and the support vectors.
   (c) Implement the method `separating_function` that takes a matrix of shape $N' \times d$ and returns a vector of size N' of evaluations of f.
   (d) Report the outputs for the code block that performs classification.