"Kernel methods in machine learning" Homework 3 Due on February 23, 2022, 3pm

Michael Arbel, Julien Mairal, and Jean-Philippe Vert

Exercice 1. 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 providing the details of the calculations:
   (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 $(\alpha_i)_{i\in\{1,\ldots,N\}}$ 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 each code block that performs a classification.

Exercice 2. Kernel Support Vector Regression

Given a dataset of N pairs (x_i, y_i), where x_i is a vector of dimension d and y_i is a scalar and an RKHS $\mathcal{H}$ of kernel k, the Kernel Support Vector Regression (Kernel SVR) finds a regression function $f \in \mathcal{H}$ and scalar b such that $f(x_i) + b - y_i$ are within and tube of size $\eta > 0$ with some tolerance. More precisely, the Kernel SVR solves the problem:

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

1. Without providing the details of the calculations:
   (a) Provide an expression for the Lagrangian of the problems in eq. (2) in terms of:
   - $2N$ dual parameters $(\alpha_i^+)_{1\leq i\leq N} \geq 0$ and $(\alpha_i^-)_{1\leq i\leq N} \geq 0$ corresponding the tube inequalities $y_i - f(x_i) - b \leq \eta + \xi_i^+$ and $-y_i + f(x_i) + b \leq \eta + \xi_i^-$

- N dual parameters μ_i^+ and μ_i^- corresponding to the positivity constraints on ξ_i^+ and ξ_i^-.

(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 $(\alpha_i^-)_{1\leq i\leq N}$, then provide an expression for $f(x)$ in terms of $(\alpha_i^+)_{1\leq i\leq N}$, $(\alpha_i^-)_{1\leq i\leq N}$ and relevant quantities.

(c) Using Strong duality, find a condition characterizing the **support vector** points x_i that are on the boundary of the tube, i.e. the points satisfying the equation $y_i - f(x_i) - b = \eta$ or $-y_i + f(x_i) + b = \eta$.

2. In the notebook, the class `KernelSVR` corresponds to eq. (2):
   (a) Implement the method `fit` that computes the optimal dual parameters α_i^+, α_i^-, the parameter, b and the support vectors.
   (b) Implement the method `regression_function` that takes a matrix of shape $M\times d$ and returns a vector of size M of evaluations of f.
   (c) Report the output of the code block that performs the regression.