Fachgebiet Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 13

We consider a class optimization problems of the type:

$$\min_{\theta} J(\theta) \quad \text{s.t.} \quad \forall_{i=1}^{m}: \; g_i(\theta) = 0 \quad \text{and} \quad \forall_{i=1}^{l}: \; h_i(\theta) \leq 0$$

For this class of problem, we can build the Lagrangian:

$$\mathcal{L}(\theta, \beta, \lambda) = J(\theta) + \sum_{i=1}^{m} \beta_i g_i(\theta) + \sum_{i=1}^{l} \lambda_i h_i(\theta).$$

where $(\beta_i)_i$ and $(\lambda_i)_i$ are the dual variables. According to the Karush-Kuhn-Tucker (KKT) conditions, it is necessary for a solution of this optimization problem that the following constraints are satisfied (in addition to the original constraints of the optimization problem):

$$\frac{\partial \mathcal{L}}{\partial \theta} = 0 \qquad \text{(stationarity)}$$

$$\forall_{i=1}^{l}: \; \lambda_i \geq 0 \qquad \text{(dual feasibility)}$$

$$\forall_{i=1}^{l}: \; \lambda_i h_i(\theta) = 0 \qquad \text{(complementary slackness)}$$

We will make use of these conditions to derive the dual form of the kernel ridge regression problem.

Exercise 1: Kernel Ridge Regression with Lagrange Multipliers (10 + 20 + 10 + 10 P)

Let $x_1, \dots, x_N \in \mathbb{R}^d$ be a dataset with labels $y_1, \dots, y_N \in \mathbb{R}$. Consider the regression model $f(x) = w^\top \phi(x)$ where $\phi \colon \mathbb{R}^d \to \mathbb{R}^h$ is a feature map and w is obtained by solving the constrained optimization problem

$$\min_{\xi, w} \; \sum_{i=1}^{N} \frac{1}{2} \xi_i^2 \quad \text{s.t.} \quad \forall_{i=1}^{N}: \; \xi_i = w^\top \phi(x_i) - y_i \quad \text{and} \quad \frac{1}{2}\|w\|^2 \leq C.$$

where equality constraints define the errors of the model, where the objective function penalizes these errors, and where the inequality constraint imposes a regularization on the parameters of the model.

(a) *Construct* the Lagrangian and *state* the KKT conditions for this problem *(Hint: rewrite the equality constraint as $\xi_i - w^\top \phi(x_i) + y_i = 0$.)*

(b) *Show* that the solution of the kernel regression problem above, expressed in terms of the dual variables $(\beta_i)_i$, and λ is given by:

$$\beta = (K + \lambda I)^{-1} \lambda y$$

where K is the kernel Gram matrix.

(c) *Express* the prediction $f(x) = w^\top \phi(x)$ in terms of the parameters of the dual.

(d) *Explain* how the new parameter λ can be related to the parameter C of the original formulation.

Exercise 2: Programming (50 P)

Download the programming files on ISIS and follow the instructions.