Assignment 3 : CS-E4830 Kernel Methods in Machine Learning 2022

The **deadline** for this assignment is **Wednesday 04.05.2022 at 4pm**. If you have **questions** about the assignment, you can ask them in the corresponding Zulip stream. We will have an exercise session regrading this assignment on 28.04.22 at 4:15 pm in TU1 Saab Auditorium.

Please follow the **submission instructions** corresponding to this assignment as given in MyCourses `https://mycourses.aalto.fi/course/view.php?id=32426&section=1`

Pen & Paper

Convex Functions

Task 1: 2 points

Recall from Lecture 7, the definition of a convex function. A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex if (i) the domain $\mathcal{X} \subseteq \mathbb{R}^n$ of f is a convex set and (ii) for all $x, y \in \mathcal{X}$, and $0 \leq \theta \leq 1$, we have

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y).$$

Also, recall the definition of the norm function from the 1st lecture. A norm on $\mathbb{R}^n$ is a function (denoted as $||.||$)

$$||.|| : \mathbb{R}^n \to \mathbb{R}^+$$

that satisfies the following requirements :

- $||v + w|| \leq ||v|| + ||w||, \forall v, w \in \mathbb{R}^n$ (Triangle Inequality)
- $||\alpha v|| = |\alpha| \times ||v||, \forall v \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$
- $||v|| \geq 0, \forall v \in \mathbb{R}^n$, and $||v|| = 0$ if and only if $v = \mathbf{0}$ (Non-negativity)

Prove that the norm function $||.||$ defined as above is a convex function on $\mathbb{R}^n$.

Task 2: 2 points

Recall from Lecture 5, the definition of a convex set. A set C is convex if

$$\forall x_1, x_2 \in C \text{ and } 0 \leq \theta \leq 1 \Rightarrow \theta x_1 + (1-\theta)x_2 \in C$$

Assuming a set C is convex (i.e., it satisfies the above definition). Then prove that, For points $x_1, x_2, x_3 \in C$ and $\theta_1, \theta_2, \theta_3 \geq 0$ such that $\theta_1 + \theta_2 + \theta_3 = 1$, the following holds

$$\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \in C$$

Task 3 - Dual of the Support Vector Machine, the C-SVM

In Lecture 6, you can see the derivation of the dual SVM, where the primal form is built on the Representer theorem. There are other primal forms of the SVM problem (such as in book by Chris Bishop: "Pattern Recognition and Machine Learning"). One of them is the so called C-SVM where the decision function is given by $f(x) = w^T\phi(x) + b$. The primal form of the soft margin C-SVM with bias term can be formulated by this optimization problem

$$\begin{array}{ll} \min_{w,\xi,b} & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i \\ \text{s.t.} & y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0,\ i = 1, \dots, m. \end{array} \tag{1}$$

Task 3a: 0.5 point

Write up the corresponding Lagrangian functional.

Task 3b: 1 point

Write up the partial derivatives of the Lagrangian functional, and derive the Karush-Kuhn-Tucker conditions connecting the primal variables to the Lagrangian dual variables.

Task 3c: 1.5 point

Finally write up the dual form of the C-SVM.

Programming

(8 points in total)

Solve the computer exercise in JupyterHub (`https://jupyter.cs.aalto.fi`). The instructions for that are given in MyCourses: `https://mycourses.aalto.fi/course/view.php?id=32426&section=4`.