CS-E4830 - Kernel Methods in Machine Learning D
Exercise 2

Alex Herrero Pons

Autumn Term Course 2020-2021

Convex Functions

Question 1: 2 points:

Recall from Lecture 7, the definition of a convex function. A function $f : \mathbb{R}^n \to \mathbb{R}$ 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 $\alpha \in \mathbb{R}$
- $||v|| \leq 0,\ \forall v \in \mathbb{R}^n$, and $||v|| = 0$ if and only if $v = 0$ (Non-negativity)

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

Solution.
First let $g(x) : \mathbb{R}^n \to \mathbb{R}$ be a norm function on $\mathbb{R}^n$.

- Then with Triangle Inequality ($||x + y|| \leq ||x|| + ||y||, \forall x, y \in \mathbb{R}^n$) we can show that:
$$g(\theta x + (1 - \theta)y) \leq g(\theta x) + g((1 - \theta)y)$$
- We also know that $||\alpha x|| = |\alpha| \cdot ||x||, \forall x \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ therefore:
$$g(\theta x) + g((1 - \theta)y) = |\theta| g(x) + |1 - \theta| g(y)$$
- And applying the non-negativity property we have that:
$$|\theta| g(x) + |1 - \theta| g(y) = (\theta)g(x) + (1 - \theta)g(y)$$

Therefore we know that $||.||$ is a convex function:

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

$\square$

Question 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

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

Solution.
Given the definition of a convex set we know that all the points in the segment $x_1 - x_2$ where $x_1, x_2 \in C$ also belong to the set C. This applies also for more points, meaning that all the points inside the figure formed by points belonging to C will also belong to C.
In this case we have three points $x_1, x_2, x_3 \in C$ that will represent a triangle (or a segment) and by giving $\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \in C$ where $\theta_1, \theta_2, \theta_3 \geq 0$ such that $\theta_1 + \theta_2 + \theta_3 = 1$ we are saying that all the points inside this triangle also belong to the set C.

□

Question 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 $f(x) = w^\top \phi(x) + b$. The primal form of the soft margin C-SVM with bias term can be formulated by this optimization problem

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

Question 3a: 0.5 point

Write up the corresponding Lagrangian functional.

Solution.

$$L(w, \xi, b, \alpha, \beta) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{m} \xi_i - \sum_{i=1}^{m} \alpha_i y_i(w^\top \phi(x_i) + b) + \sum_{i=1}^{m} \alpha_i(1 - \xi_i) - \sum_{i=1}^{m} \beta_i \xi_i$$

Question 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.

Solution.
The corresponding partial derivatives are:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{m} \alpha_i y_i \phi(x_i)$$

$$\frac{\partial L}{\partial \xi} = C - \alpha_i - \beta_i$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{m} \alpha_i y_i$$

And the Karush-Kuhn-Tucker conditions:

$$\frac{\partial L}{\partial w} = 0; \quad \frac{\partial L}{\partial \xi} = 0; \quad \frac{\partial L}{\partial b} = 0;$$

$$w = \sum_{i=1}^{m} \alpha_i y_i \phi(x_i); \quad \beta_i \geq 0; \quad 0 \leq \alpha_i \leq C; \quad \sum_{i=1}^{m} \alpha_i \beta_i = 0;$$

Therefore the Lagrangian from the previous subsection (3a) can be rewritten as:

$$L(w, \xi, b, \alpha, \beta) = \frac{1}{2}||\sum_{i=1}^{m} \alpha_i y_i \phi(x_i)||^2 + C\sum_{i=1}^{m} \xi_i - \sum_{i=1}^{m} \alpha_i y_i (\phi(x_i)^\top \sum_{i=1}^{m} \alpha_i y_i \phi(x_i) + b) +$$

$$+ \sum_{i=0}^{m} \alpha_i (1 - \xi_i) - \sum_{i=1}^{m} \beta_i \xi_i =$$

$$= \frac{1}{2}||\sum_{i=1}^{m} \alpha_i y_i \phi(x_i)||^2 - \sum_{i=1}^{m} \alpha_i y_i (\phi(x_i)^\top \sum_{i=1}^{m} \alpha_i y_i \phi(x_i) + b) =$$

$$= \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j)$$

Question 3c: 1.5 point

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

Solution.

$$\max_{\alpha} L(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j)$$