Assignment 3 : CS-E4830 Kernel Methods in Machine Learning 2019

The **deadline** for this assignment is **Thursday 28.03.2019 at 4pm**. If you have **questions** about the assignment, you can ask them in the 'General discussion' section on MyCourses. We will have a session regarding the **solutions** of this assignment on 28.03.19 at 4:15 pm in TU1(1017), TUAS, Maarintie 8. The solutions will also be available in MyCourses.

Please follow the **submission instructions** given in MyCourses: `https://mycourses.aalto.fi/course/view.php?id=20602&section=2`.

Important : In solving any of the below exercises, if you are referring some book or resource on internet or discussing with someone, please write the reference to the source explicitly along with that exercise.

Pen & Paper exercise

Convex Functions

Question 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 $||.||$)

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

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 $\mathbb{R}^n$.

Solution

We have the function $f(x) = ||x||$, thus that we need to prove is this statement

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

Based on the properties of the norm function the left hand side can be transformed step by step. By the first rule

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

By the second rule, since $0 \le \theta \le 1$

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

By combining these two expressions we have the statement of the question.

Dual of the Support Vector Machine, the C-SVM

In Lecture 8, 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

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

Question 2: 0.5 point

Write up the corresponding Lagrangian functional.

Question 3: 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.

Question 4: 1.5 point

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

Solution

Let the Lagrangian α_i be assigned to the constraint $y_i(w^T\phi(x_i)+b) \ge 1-\xi_i$ for all i, and β_i be assigned to $\xi_i \ge 0$ for all i as well.

Since the constraints are given as inequalities we have

$$\alpha_i \ge 0, \quad \beta_i \ge 0$$

for all i.

Let the constraints be rewritten as

$$y_i(w^T\phi(x_i)+b) \ge 1-\xi_i \Rightarrow 1-\xi_i - y_i(w^T\phi(x_i)+b) \le 0$$

and $\xi_i \ge 0 \Rightarrow -\xi \le 0$ for all i.

The Lagrangian functional takes this form

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

The partial derivatives with respect to the primal variables

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{m} \alpha_i y_i \phi(x_i) = 0, \ \Rightarrow \boxed{\mathbf{w} = \sum_{i=1}^{m} \alpha_i y_i \phi(x_i)}, \tag{2}$$

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial b} = \boxed{\sum_{i=1}^{m} \alpha_i y_i = 0}. \tag{3}$$

Note b can not be expressed as function of the dual variables.

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \xi_i} = C - \alpha_i - \beta_i = 0 \text{ for all } i. \tag{4}$$

Since $\beta_i \geq 0$, and $\alpha_i \geq 0$ we have $\boxed{0 \leq \alpha_i \leq C}$ for all i. ξ_i can not be written as function of the dual variables.

We have the constrains for the dual variables:

$\boxed{\sum_{i=1}^{m} \alpha_i y_i = 0}$ and $\boxed{0 \leq \alpha_i \leq C}$, $i = 1, \dots, m$.

To derive the dual we need to substitute the expression of $\mathbf{w}$ back into the Lagrangian. First the Lagrangian functional is restructured

$$\begin{aligned} L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \alpha_i(1 - \xi_i - y_i(w^T\phi(x_i) + b)) - \sum_{i=1}^{m} \beta_i \xi_i \\ &= \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m} \alpha_i \xi_i - \sum_{i=1}^{m} \alpha_i y_i w^T\phi(x_i) \\ &\quad - \sum_{i=1}^{m} \alpha_i y_i b - \sum_{i=1}^{m} \beta_i \xi_i \\ &= \tfrac{1}{2}\|w\|^2 + \sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m} \alpha_i y_i w^T\phi(x_i) \\ &\quad \boxed{+C\sum_{i=1}^{m} \xi_i - \sum_{i=1}^{m} \alpha_i \xi_i - \sum_{i=1}^{m} \beta_i \xi_i} - b\boxed{\sum_{i=1}^{m} \alpha_i y_i} \end{aligned}$$

The terms in the boxes are equal to 0 based on Expressions (4) and (3). After eliminating those term we can make the substitution

$$\begin{aligned} L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \tfrac{1}{2}\|w\|^2 + \sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m} \alpha_i y_i w^T\phi(x_i) \\ &= \tfrac{1}{2}\sum_{i=1}^{m} \alpha_i y_i \phi(x_i)^T \sum_{j=1}^{m} \alpha_j y_j \phi(x_j) + \sum_{i=1}^{m} \alpha_i \\ &\quad - \sum_{i=1}^{m} \alpha_i y_i \sum_{j=1}^{m} \alpha_j y_j \phi(x_j)^T \phi(x_i) \\ &= \tfrac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j) + \sum_{i=1}^{m} \alpha_i \\ &\quad - \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j) \\ &= \boxed{-\tfrac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j) + \sum_{i=1}^{m} \alpha_i}, \end{aligned}$$

where we used the kernel expression $\kappa(x_i, x_j) = \phi(x_i)^T\phi(x_j)$.

Finally we write up the dual

$$\begin{array}{ll} \text{max} & -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \kappa(x_i, x_j) + \sum_{i=1}^{m} \alpha_i \\ \text{with respect to} & \alpha_i, \ i = 1, \dots, m \\ \text{subject to} & \sum_{i=1}^{m} \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, \ i = 1, \dots, m. \end{array}$$

Question for Bonus points - Convex Sets

Question 5: 1 point

Recall from Lecture 7, 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 Let assume that $\theta \neq 1$ then we can write

$$\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 = \theta_1 x_1 + \frac{1-\theta_1}{1-\theta_1}\theta_2 x_2 + \frac{1-\theta_1}{1-\theta_1}\theta_3 x_3.$$

Let the right hand side rearranged

$$\theta_1 x_1 + \frac{1-\theta_1}{1-\theta_1}\theta_2 x_2 + \frac{1-\theta_1}{1-\theta_1}\theta_3 x_3 = \theta_1 x_1 + (1-\theta_1)\underbrace{\left(\frac{\theta_2}{1-\theta_1}x_2 + \frac{\theta_3}{1-\theta_1}x_3\right)}_{=x}$$

$x \in C$ because $x_2, x_3 \in C$, and

$$\frac{\theta_2}{1-\theta_1} + \frac{\theta_3}{1-\theta_1} = \frac{\theta_2+\theta_3}{1-\theta_1} = \frac{\theta_2+\theta_3}{\theta_2+\theta_3} = 1,$$

since $\theta_1 + \theta_2 + \theta_3 = 1$.

Now we have $x_1, x \in C$, and $\theta_1 + (1-\theta_1) = 1$, thus $\theta_1 x_1 + (1-\theta_1)x \in C$