CS-E4830 - Kernel Methods in Machine Learning

Homework Assignment 3 – Pen and Paper

Student: Quoc Tuan Vinh, Ngo (704526)

Convex Functions

Question 1: *Prove that the norm function* $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}^+$ *is a convex function on* $\mathbb{R}^n$

We learn that a norm on $\mathbb{R}^n$ satisfies the following requirement:

$\forall v, w \in \mathbb{R}^n, \|v + w\| \leq \|v\| + \|w\|$ (Triangle Inequality)

$\forall v \in \mathbb{R}^n, \forall \alpha \in \mathbb{R}, \|\alpha v\| = |\alpha| \times \|v\|$ (Absolute Scalability)

$\forall v \in \mathbb{R}^n, \|v\| \geq 0, and\ \|v\| = 0\ if\ and\ only\ if\ v = \mathbf{0}$ (Non-negativity)

To prove norm is a convex function $f: \mathbb{R}^n \rightarrow \mathbb{R}^+$, we need to prove:

(i) Domain $\mathcal{X} \subseteq \mathbb{R}^n$ of f is a convex set
(ii) $\forall v, w \in \mathbb{R}^n$ and $\theta \in \mathbb{R},\ 0 \leq \theta \leq 1$, we have:

$$f(\theta v + (1-\theta)w) \leq \theta f(v) + (1-\theta) f(w)$$

Let $f: \mathbb{R}^n \rightarrow \mathbb{R}^+$ denote the norm function. $\forall v, w \in \mathbb{R}^n$ and $\theta \in \mathbb{R},\ 0 \leq \theta \leq 1$, we can derive the following expansion:

$f(\theta v + (1-\theta)w)$

$\leq f(\theta v) + f\big((1-\theta)w\big)$ (due to Triangle Inequality)

$= |\theta| \times f(v) + |1-\theta| f(w)$ (due to Absolute Scalability)

$= \theta \times f(v) + (1-\theta) f(w)$ (because $0 \leq \theta \leq 1$)

Which satisfies (ii).

And since we define $\forall v, w \in \mathbb{R}^n$, (i) is also satisfied. Therefore, we can conclude that norm function $\|\cdot\|$ defined as above is a convex function on $\mathbb{R}^n$.

Dual of the Support Vector Machine, the C-SVM

Optimization problem:

$$min_{\boldsymbol{w},\boldsymbol{\xi},\boldsymbol{b}} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i$$

$$\text{s.t. } y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0, i = 1, \dots, m$$

Question 2: *Write up Lagrangian functional*

$$\mathcal{L}(\boldsymbol{w}, \boldsymbol{\xi}, b, \boldsymbol{\alpha}, \boldsymbol{v}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \alpha_i[1 - \xi_i - y_i(\boldsymbol{w}^{\boldsymbol{T}}\phi(x_i) + b)] - \sum_{i=1}^{m} v_i \xi_i$$

Question 3: *Write up the partial derivatives of Lagrangian functional*

$\mathcal{L}(\boldsymbol{w}, \boldsymbol{\xi}, b, \boldsymbol{\alpha}, \boldsymbol{v})$ is a convex quadratic function in $\boldsymbol{w}$**.** To find optimal value, we should derive it with restrict to $\boldsymbol{w}, \boldsymbol{\xi}, b$ and set the gradient to **0.**

(i) Derivatives wrt to $\boldsymbol{w}$**:**

$$\nabla_w \mathcal{L} = \frac{\delta \mathcal{L}}{\delta \boldsymbol{w}} = \boldsymbol{w} + \sum_{i=1}^{m} \alpha_i(-y_i \phi(x_i))$$

Set $\nabla_w \mathcal{L} = 0$, we have: $w = \sum_{i=1}^{m} \alpha_i(y_i \phi(x_i))$ **(1)**

(ii) Derivatives wrt to ξ_i**:**

$$\nabla_{\xi_i} \mathcal{L} = \frac{\delta \mathcal{L}}{\delta \xi_i} = C - \alpha_i - v_i$$

Set $\nabla_{\xi_i} \mathcal{L} = 0$, we have: $C = \alpha_i + v_i$ **(2)**

(iii) Derivatives wrt to b:

$$\nabla_{\boldsymbol{b}} \mathcal{L} = \frac{\delta \mathcal{L}}{\delta b} = -\sum_{i=1}^{m} \alpha_i y_i$$

Set $\nabla_{\boldsymbol{b}} \mathcal{L} = 0$, we have $\sum_{i=1}^{m} \alpha_i y_i = 0$ **(3)**

Write up the KKT conditions: $\forall i = 1, \dots, m$:

$\alpha_{\boldsymbol{i}}, v_{\boldsymbol{i}}, \xi_{\boldsymbol{i}} \geq 0$

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

$\alpha_i[1 - \xi_i - y_i(w^T \phi(x_i) + b)] = 0$

$v_{\boldsymbol{i}} \xi_{\boldsymbol{i}} = 0$

$$\boldsymbol{w} = \sum_{i=1}^{m} \alpha_i(y_i \phi(x_i))$$

$C = \alpha_i + v_i$

$$\sum_{i=1}^{m} \alpha_i y_i = 0$$

Question 4. *Write up the dual form of the C-SVM*

Replace the results of (1), (2), (3) into our optimization problem, we can obtain the dual form of the C-SVM as follows:

$$max_{\alpha_i} -\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sum_{i=1}^{m} \alpha_i$$

such that $\forall i = 1, \dots, m$: $0 \leq \alpha_i \leq C$, since $C = \alpha_i + v_i$, and $\sum_{i=1}^{m} \alpha_i y_i = 0$.

In the function above:

- $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is simply the kernel matrix between $\phi(\boldsymbol{x}_i)$ *and* $\phi(\boldsymbol{x}_j)$
- $\boldsymbol{\alpha}, \boldsymbol{v} \in \mathbb{R}^{\boldsymbol{m}}$ are the dual variable vectors.
- $\boldsymbol{y}$ is the training label.
- $C > 0 \in \mathbb{R}$ is the regularization parameter

Convex Sets

Question 5. *Prove the convex combination*

Recall from Lecture 7, a set C is convex if

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

$\forall x_1, x_2, x_3 \in C, and\ 0 \leq \theta \leq 1, and\ \theta_1 + \theta_2 + \theta_3 = 1$:

(i) If $\theta_1 = 1$, meaning that $\theta_2 = \theta_3 = 0$. In this case, the proof is automatically provided as $\forall x_1 \in C$ by definition. **(2)**

(ii) If $\theta_1 \neq 1$: $\theta_2 + \theta_3 = 1 - \theta_1$ $(\neq 0)$

$$\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 = \theta_1 x_1 + (1-\theta_1)\left[\frac{\theta_2 x_2}{(1-\theta_1)} + \frac{\theta_3 x_3}{(1-\theta_1)}\right]$$
$$= \theta_1 x_1 + (1-\theta_1)\left(\frac{\theta_2 x_2}{\theta_2 + \theta_3} + \frac{\theta_3 x_3}{\theta_2 + \theta_3}\right) = \theta_1 x_1 + (1-\theta_1)(\theta_2^* x_2 + \theta_3^* x_3)$$

where $\theta_2^* = \frac{\theta_2}{\theta_2 + \theta_3}$ and $\theta_3^* = \frac{\theta_3}{\theta_2 + \theta_3}$. It can be seen that $\theta_2^* + \theta_3^* = 1, and\ 0 \leq \theta_2^*, \theta_3^* \leq 1$. Since $\forall x_2, x_3 \in C$, we can conclude that $y = (\theta_2^* x_2 + \theta_3^* x_3) \in C$.

Therefore, we can have that:

$\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 = \theta_1 x_1 + (1-\theta_1)y$ also $\in C$. **(3)**

From (2), (3), we can have the required proof.