Kernel Methods in Machine Learning
Homework Assignment 3

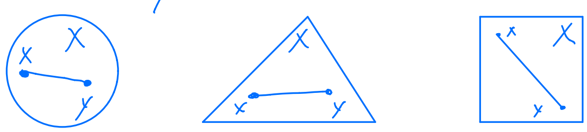
student : Donayorov Maksad
id : 571335

Question 1:

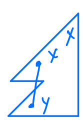
$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) \cdot f(y)$

$f(x)$ is considered to be a convex function if $x, y \in X$ and X is a convex set and $0 \leq \theta \leq 1$.

Convex set representation:

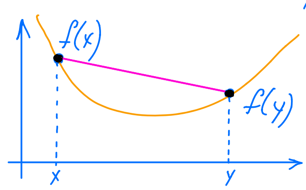


$\Rightarrow X$ is convex set



$\Rightarrow X$ is not convex set

Convex function representation.



$f(\theta x + (1-\theta) \cdot y)$ – function value

$\theta \cdot f(x) + (1-\theta) \cdot f(y)$ – interpolation

Let's asume that $g(x) : \mathbb{R}^n \to \mathbb{R}$ is a norm, then:

1. triangular inequality $\|x+y\| \leq \|x\| + \|y\|$, $\forall x, y \in \mathbb{R}^n$
   holds for $g(\theta x + (1-\theta)y) \leq \theta \cdot g(x) + (1-\theta) \cdot g(x)$
2. $\|\alpha \cdot x\| = |\alpha| \cdot \|x\|$, $\forall x \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ holds since
   $|\theta| \cdot g(x) + |1-\theta| \cdot g(y)$
3. Non-negativity holds as:
   $|\theta| \cdot g(x) + |1-\theta| \cdot g(y) = \theta \cdot g(x) + (1-\theta) \cdot g(y)$

Consequently, $\|.\|$ is convex.

Question 2:

$$\min_{w,\varepsilon,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \varepsilon_i$$
$$y_i(w^T\phi(x_i)+b) \geq 1-\varepsilon_i$$
$$\varepsilon_i \geq 0,\ i = 0 \dots m$$

corresponding Lagrangian:

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

assuming that $\delta_i, \gamma_i \geq 0$

Question 3:

partial derivative withe respect to w:

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

partial derivative withe respect to ε_i:

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

partial derivative withe respect to b:

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

Karush-Kuhn-Tucker condition:

$$\frac{\partial \mathcal{L}}{\partial w} = 0 \qquad \frac{\partial \mathcal{L}}{\partial \varepsilon} = 0 \qquad \frac{\partial \mathcal{L}}{\partial b} = 0$$

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

The Lagrangian can be written as:

$$\mathcal{L} = \frac{1}{2}\left\|\sum_{i=1}^{m} \alpha_i y_i \phi(x_i)\right\|^2 + C\sum_{i=1}^{m} \varepsilon_i - \sum_{i=1}^{m} \alpha_i y_i \left(\phi(x_i)^T \sum_{j=1}^{m} \alpha_j y_j \phi(x_j) + b\right) + \sum_{i=1}^{m} \alpha_i(1-\varepsilon_i) - \sum_{i=1}^{m} \beta_i \varepsilon_i$$

$= 0$

and the Lagrangian can be rewritten as:

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

Question 4:

Having:

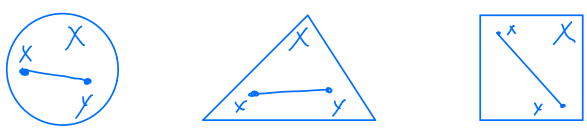
$$\mathcal{L} = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i y_i \alpha_j y_j K(x_i, x_j)$$

the dual problem is

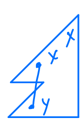
$$\max_{\alpha} \mathcal{L}(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i y_i \alpha_j y_j K(x_i, x_j)$$

Question 5:

Similar to Question 1 convex set can be represented such:



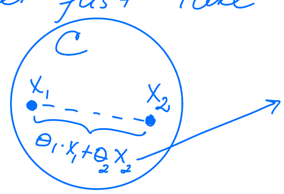
$\Rightarrow$ X is convex set



$\Rightarrow$ X is not convex set

To prove that $\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 \in C$ such that $x_1, x_2, x_3 \in C$ and $\theta_1, \theta_2, \theta_3 \geq 0$ $\theta_1 + \theta_2 + \theta_3 = 1$;

let just take 2 points only (x_1, x_2)



this implies that θ_1 percent of x_1 and θ_2 percent of x_2. For example if $\theta_1 = 0.3$ then $\theta_2 = 0.7$ and $0.3 \cdot x_1 + 0.7 \cdot x_2$ will be $\in C$.

and adding point x_3 will have θ_3 such that $\theta_1 + \theta_2 + \theta_3 = 1$ which means $\theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \theta_3 \cdot x_3$ will be $\in C$.