AI 501 Machine Learning for AI, Spring 2021

# Homework assignment 3

1. (15 points) A set $\boldsymbol{X} \subset \mathbb{R}^n$ is convex if for any $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X}$ and $t \in [0, 1]$, we have

$$t\boldsymbol{x} + (1-t)\boldsymbol{y} \in \boldsymbol{X}. \tag{1}$$

   Using this definition of the convex set, state 'True' or 'False' that following sets are convex (Hint: you don't need to justify your answer for this problem — just write down True/False for each question):

   (a) (3 points) $\boldsymbol{X} = \varnothing$ (i.e., an empty set).

   (b) (3 points) $\boldsymbol{X} = \mathbb{R}^n$ (i.e., the entire space).

   (c) (3 points) For some $r \in \mathbb{R}$, $\boldsymbol{X} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x}^\top \boldsymbol{x} \le r\}$.

   (d) (3 points) For some $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{X} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \forall t \in [0, \infty), \boldsymbol{x} = t\boldsymbol{v}\}$.

   (e) (3 points) For some $\boldsymbol{a} \in \mathbb{R}^n$, $\boldsymbol{X} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}^\top \boldsymbol{x} = 0\}$.

2. (25 points) A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\mathrm{dom}(f)$ is a convex set and if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}(f)$ and $t \in [0, 1]$, we have

$$f(t\boldsymbol{x} + (1-t)\boldsymbol{y}) \le tf(\boldsymbol{x}) + (1-t)f(\boldsymbol{y}). \tag{2}$$

   Using this definition of the convex function, show that following functions are convex:

   (a) (7 points) $\max$ function on $\mathbb{R}^n$:

$$f(\boldsymbol{x}) = f(x_1, ..., x_n) = \max\{x_1, ..., x_n\}. \tag{3}$$

   (b) (7 points) $\ell_p$ norms on $\mathbb{R}^n$:

$$f(\boldsymbol{x}) = f(x_1, ..., x_n) = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}. \tag{4}$$

   (c) (11 points) $\mathrm{LogSumExp}$ function on $\mathbb{R}^n$ (Hint: you can use the Hölder's inequality):

$$f(\boldsymbol{x}) = f(x_1, ..., x_n) = \log\left(\sum_{i=1}^{n} e^{x_i}\right). \tag{5}$$

3. (40 points) Convexity Conditions.

   (a) (20 points) Show that a differentiable function $f : \mathbb{R} \to \mathbb{R}$ is convex iff the following inequality holds:

$$\forall x, y \in \mathrm{dom}(f), \quad f(y) \ge f(x) + f'(x)(y - x). \tag{6}$$

   (b) (20 points) Show that a twice-differentiable function $f : \mathbb{R} \to \mathbb{R}$ is convex iff the following inequality holds (Hint: you can use the above condition, i.e., 6, to solve this problem):

$$\forall z \in \mathrm{dom}(f), \quad f''(z) \ge 0. \tag{7}$$

4. (20 points) Suppose that we have linearly separable dataset

$$\mathcal{D} = \{\boldsymbol{x}_i, y_i\}_{i=1}^{N}, \tag{8}$$

which satisfies

$$y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq 1, \quad \text{for } i = 1, ..., N \tag{9}$$

for some $\boldsymbol{w}$ and b, where $y_i \in \{-1, 1\}$ denotes a class of $\boldsymbol{x}_i \in \mathbb{R}^n$. Then, we can formulate the primal problem for a **hard-margin support vector machine** as follows:

$$\begin{aligned} \underset{\boldsymbol{w},\, b}{\text{minimize}} \quad & \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{w} \\ \text{subject to} \quad & y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq 1, \quad \text{for } i = 1, ..., N. \end{aligned} \tag{10}$$

From this, you should derive the Lagrangian dual problem:

$$\begin{aligned} \underset{\boldsymbol{\lambda}}{\text{maximize}} \quad & \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \boldsymbol{x}_i^\top \boldsymbol{x}_j \\ \text{subject to} \quad & \lambda_i \geq 0, \quad \text{for } i = 1, ..., N, \\ & \sum_{i=1}^{N} \lambda_i y_i = 0. \end{aligned} \tag{11}$$

To do this, you should (i) define the *Lagrangian*, i.e., $\mathcal{L}(\boldsymbol{w}, b, \boldsymbol{\lambda})$ from 10, and (ii) define the *Lagrangian dual function* defined by the infimum value of the *Lagrangian*, i.e., $\inf_{\boldsymbol{w},b} \mathcal{L}(\boldsymbol{w}, b, \boldsymbol{\lambda})$. (iii) Does the strong duality holds in this situation? Justify your answer.