Tsinghua-Berkeley Shenzhen Institute
LEARNING FROM DATA
Fall 2018

Problem Set 1

Issued: Monday 8th October, 2018 **Due:** Monday 15th October, 2018

POLICIES

- **Acknowledgments:** We expect you to make an honest effort to solve the problems individually. As we sometimes reuse problem set questions from previous years, covered by papers and web pages, we expect the students **NOT** to copy, refer to, or look at the solutions in preparing their answers (relating to an unauthorized material is considered a violation of the honor principle). Similarly, we expect to not to google directly for answers (though you are free to google for knowledge about the topic). If you do happen to use other material, it must be acknowledged here, with a citation on the submitted solution.

- **Required homework submission format:** You can submit homework either as one single PDF document or as handwritten papers. Written homework needs to be provided during the class in the due date, and PDF document needs to be submitted through Tsinghua's Web Learning (`http://learn.tsinghua.edu.cn/`) before the end of due date.

  It is encouraged you LaTeX all your work, and we would provide a LaTeX template for your homework.

- **Collaborators:** In a separate section (before your answers), list the names of all people you collaborated with and for which question(s). If you did the HW entirely on your own, **PLEASE STATE THIS**. Each student must understand, write, and hand in answers of their own.

Consider the problem of classifying l samples $(\boldsymbol{x}_i, y_i)$ using SVM, where $\boldsymbol{x}_i \in \mathbb{R}^n$, $y_i \in \{-1, 1\}$, $(i = 1, \cdots, l)$.

1.1. Suppose the data are linearly separable. The optimization problem of SVM is

$$\begin{aligned} \underset{\boldsymbol{w},b}{\text{minimize}} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 \\ \text{subject to} \quad & y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1, \quad i = 1, \dots, l, \end{aligned} \tag{P}$$

and let $(\boldsymbol{w}^\star, b^\star)$ denote its optimal solution.

(a) Show that

$$b^\star = -\frac{1}{2}\left(\max_{i\colon y_i=-1} \boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i + \min_{i\colon y_i=1} \boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i\right).$$

The corresponding Lagrange dual problem is given by

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{maximize}} \quad & \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \\ \text{subject to} \quad & \alpha_i \geq 0, \quad i = 1, \dots, l, \\ & \sum_{i=1}^{l} \alpha_i y_i = 0. \end{aligned} \tag{D}$$

Suppose the optimal solution of (D) is $\boldsymbol{\alpha}^\star = (\alpha_1^\star, \cdots, \alpha_l^\star)^{\mathrm{T}}$, from the KKT conditions we know that

$$\begin{aligned} \boldsymbol{w}^\star &= \sum_{i=1}^{l} \alpha_i^\star y_i \boldsymbol{x}_i, \\ \sum_{i=1}^{l} \alpha_i^\star & \left[y_i(\boldsymbol{w}^{\star\mathrm{T}} \boldsymbol{x}_i + b^\star) - 1 \right] = 0. \end{aligned} \tag{1}$$

(b) Based on (1), verify that

$$\frac{1}{2}\|\boldsymbol{w}^\star\|_2^2 = \sum_{i=1}^{l} \alpha_i^\star - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i^\star \alpha_j^\star y_i y_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = \frac{1}{2}\sum_{i=1}^{l} \alpha_i^\star.$$

1.2. When the data are not linearly separable, consider the soft-margin SVM given by

$$\begin{aligned} \underset{\boldsymbol{w}, b, \boldsymbol{\xi}}{\text{minimize}} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{l} \xi_i \\ \text{subject to} \quad & \xi_i \geq 0, \quad i = 1, \dots, l, \\ & y_i(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, l, \end{aligned} \tag{2}$$

where $C > 0$ is a fixed parameter.

(a) Show that (2) is equivalent[1] to

$$\underset{\boldsymbol{w}, b}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{l} \ell(y_i, \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i + b), \tag{3}$$

where $\ell(\cdot, \cdot)$ is the hinge loss defined by $\ell(y, z) \triangleq \max\{1 - yz, 0\}$.

(b) Show that the objective function of (3), denoted by $f(\boldsymbol{w}, b)$, is convex, i.e.,

$$f(\theta \boldsymbol{w}_1 + (1-\theta)\boldsymbol{w}_2, \theta b_1 + (1-\theta) b_2) \leq \theta f(\boldsymbol{w}_1, b_1) + (1-\theta) f(\boldsymbol{w}_2, b_2).$$

for all $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathbb{R}^n, b_1, b_2 \in \mathbb{R}$, and $\theta \in [0, 1]$.

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge university press, 2004.

[1] Two optimization problems are called equivalent if from a solution of one, a solution of the other is readily found, and vice versa.