Tsinghua-Berkeley Shenzhen Institute
Learning from Data
Fall 2018

Homework 1

YOUR NAME — October 10, 2018

- **Acknowledgments:** This template takes some materials from course CSE 547/Stat 548 of Washington University: `https://courses.cs.washington.edu/courses/cse547/17sp/index.html`.
- **Collaborators:** I finish my homework all by myself.

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

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

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

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

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

Proof. From a condition from the (1):

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

There is at least one $\alpha_j^\star > 0$ that let $\boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_j + b^\star - y_j = 0$.
As the data are linearly separable, for any $y_j = 1$, $\boldsymbol{w}^\star\boldsymbol{x}_j + b^\star \geq 1$,
So that

$$\min_{j:y_j=1} \{\boldsymbol{w}^\star\boldsymbol{x}_j + b^\star\} = 1 \tag{1.a.1}$$

Similarily, for any $y_j = -1$, $\boldsymbol{w}^\star\boldsymbol{x}_j + b^\star \leq -1$, So that

$$\max_{j:y_j=-1} \{\boldsymbol{w}^\star\boldsymbol{x}_j + b^\star\} = -1 \tag{1.a.2}$$

Sum up the equation 1.a.1 and 1.a.2, here it is:

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

□

(b) Based on (1), verify that

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

Proof. From a condition from the (1):

$$\sum_{i=1}^{l} \alpha_i^\star \left[y_i(\boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i + b^\star) - 1 \right] = 0$$
$$\sum_{i=1}^{l} \alpha_i^\star y_i \boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i + b^\star \sum_{i=1}^{l} \alpha_i^\star y_i - \sum_{i=1}^{l} \alpha_i^\star = 0$$
$$\sum_{i=1}^{l} \alpha_i^\star = \sum_{i=1}^{l} \alpha_i^\star y_i \boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i = \boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{w}^\star = \|\boldsymbol{w}^\star\|_2^2$$

And as

$$\sum_{i=1}^{l} \alpha_i^\star y_i \boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i = \sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i^\star\alpha_j^\star y_i y_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$$

Here it is:

$$\begin{aligned} \frac{1}{2}\|\boldsymbol{w}^\star\|_2^2 &= \boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{w}^\star - \frac{1}{2}\boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{w}^\star \\ &= \sum_{i=1}^{l} \alpha_i^\star - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i^\star\alpha_j^\star y_i y_j \langle \boldsymbol{x}_i \boldsymbol{x}_j \rangle \end{aligned}$$

□

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

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

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 $\ell(y,z) \triangleq \max\{1 - yz, 0\}$.

Proof. Let $\xi_i = \ell(y_i, \boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i + b)$, As

$$\ell(y_i, \boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i + b) = \max(0, 1 - y_i(\boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i, +b))$$

The $\xi_i \geq 0, \quad i = 1,\ldots,l.$
When

$$1 - y_i(\boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i + b) > 0$$

The $\xi_i = 1 - y_i(\boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i + b)$, so that:

$$y_i(\boldsymbol{w}^{\star\mathrm{T}} + b) = 1 - \xi_i \tag{2.a.1}$$

When

$$1 - y_i(\boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i + b) \leq 0$$

The $\xi_i = 0$, so that:

$$y_i(\boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i + b) \geq 1 - 0 \geq 1 - \xi_i \tag{2.a.2}$$

Combine the (2.a.1) and (2.a.2), it is easy to acquire that

$$y_i(\boldsymbol{w}^{\star\mathrm{T}}\boldsymbol{x}_i + b) \geq 1 - \xi_i, \quad i = 1,\ldots,l.$$

Taking all above into consideration, the (3) can be transformed as:

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

Which is equivalent to the (2). □

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

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

$$f(\theta\boldsymbol{w}_1+(1-\theta)\boldsymbol{w}_2, \theta b_1+(1-\theta)b_2) \leq \theta f(\boldsymbol{w}_1, b_1)+(1-\theta)f(\boldsymbol{w}_2, b_2). \quad (4)$$

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

Proof. The left part of the (4):

$$\begin{aligned}
&f(\theta\boldsymbol{w}_1 + (1-\theta)\boldsymbol{w}_2, \theta b1 + (1-\theta)b2) \\
&= \frac{\theta^2}{2}\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_1 + \frac{(1-\theta)^2}{2}\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_2 + \frac{\theta(1-\theta)}{2}(\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_2 + \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_1) && (5a) \\
&+ C\sum_{i=1}^{l} \ell\left(y_i, (\theta\boldsymbol{w}_1 + (1-\theta)\boldsymbol{w}_2)^{\mathrm{T}}x_i + \theta b_1 + (1-\theta)b_2\right) && (5b)
\end{aligned}$$

The right part of the (4):

$$\begin{aligned}
&\theta f(\boldsymbol{w}_1, b_1) + (1-\theta)f(\boldsymbol{w}_2, b_2) \\
&= \frac{\theta}{2}\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_1 + \frac{(1-\theta)}{2}\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_2 && (6a) \\
&+ \theta C\sum_{i=1}^{l}\ell(y_i, \boldsymbol{w}_1^{\mathrm{T}}x_i + b_1) + (1-\theta)C\sum_{i=1}^{l}\ell(y_i, \boldsymbol{w}_2^{\mathrm{T}}x_i + b_2) && (6b)
\end{aligned}$$

The (6a) - (5a) is:

$$\begin{aligned}
&\frac{\theta^2}{2}(\boldsymbol{w}_1 - \boldsymbol{w}_2)^{\mathrm{T}}(\boldsymbol{w}_1 - \boldsymbol{w}_2) + \frac{\theta}{2}(\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_2 + \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_1 - \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_2) + \frac{(1-\theta)}{2}\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_2 \\
&-(\frac{\theta}{2}\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_1 + \frac{(1-\theta)}{2}\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_2) \\
=&\frac{\theta}{2}\left(\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_1 - \theta(\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_1 + \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_2 - \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_2 - \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_1) + \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_2 - \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_2 - \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_1\right) \\
=&\frac{\theta(1-\theta)}{2}(\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_1 + \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_2 - \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_2 - \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_1) \\
=&\frac{\theta(1-\theta)}{2}(\boldsymbol{w}_1 - \boldsymbol{w}_2)^{\mathrm{T}}(\boldsymbol{w}_1 - \boldsymbol{w}_2) \geq 0.
\end{aligned}$$

So that the (5a) $\leq$ (6a).
Similarly, let $\boldsymbol{z}_{1,i} = \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{x}_i + b_1$ and $\boldsymbol{z}_{2,i} = \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{x}_i + b_2$,
The (6b) - (5b) is:

$$\begin{aligned}
&\theta C\sum_{i=1}^{l}\ell(y_i, \boldsymbol{z}_{1,i}) + (1-\theta)C\sum_{i=1}^{l}\ell(y_i, \boldsymbol{z}_{2,i}) - C\sum_{i=1}^{l}\ell(y_i, \theta\boldsymbol{z}_{1,i} + (1-\theta)\boldsymbol{z}_{2,i}) \\
=&C\theta\sum_{i=1}^{l}\max\{0, 1 - y_i\boldsymbol{z}_{1,i}\} + C(1-\theta)\sum_{i=1}^{l}\max\{0, 1 - y_i\boldsymbol{z}_{2,i}\} && (7a) \\
&- C\sum_{i=1}^{l}\max\{0, 1 - y_i(\theta\boldsymbol{z}_{1,i} + (1-\theta)\boldsymbol{z}_{2,i})\} && (7b)
\end{aligned}$$

Since $\max\{u, v\} = \frac{1}{2}(u + v + |u - v|)$,

$$(7\text{a}) = C \sum_{i=1}^{l} [1 - y_i(\theta \boldsymbol{z}_{1,i} + (1-\theta)\boldsymbol{z}_{2,i}) + |1 - y_i(\theta \boldsymbol{z}_{1,i} + (1-\theta)\boldsymbol{z}_{2,i})|]$$

$$(7\text{b}) = C \sum_{i=1}^{l} [1 - \theta y_i \boldsymbol{z}_{1,i} - (1-\theta) y_i \boldsymbol{z}_{2,i} + |\theta(1 - y_i \boldsymbol{z}_{1,i})| + |(1-\theta)(1 - y_i \boldsymbol{z}_{2,i})|]$$

The (6b) - (5b) is:

$$\begin{aligned}(7\text{b}) - (7\text{a}) = C \sum_{i=1}^{l} [&|\theta(1 - y_i \boldsymbol{z}_{1,i})| + |(1-\theta)(1 - y_i \boldsymbol{z}_{2,i})| \\ &- |\theta(1 - y_i \boldsymbol{z}_{1,i}) + (1-\theta)(1 - y_i \boldsymbol{z}_{2,i})| \geq 0\end{aligned}$$

That is the (5b) $\leq$ (6b).
With all above, the (5a) + (5b) $\leq$ (6a) + (6b), so the (4) is proved to be true. $\square$