Tsinghua-Berkeley Shenzhen Institute
Learning from Data
Fall 2019

Writing Homework 2

TIAN Chenyu October 21, 2019

- **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 this homework by myself.

2.1. Define $\boldsymbol{P} = \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}$, for a given vector $\boldsymbol{v}$

$$\boldsymbol{v} = \boldsymbol{P}\boldsymbol{v} + (\boldsymbol{v} - \boldsymbol{P}\boldsymbol{v})$$

If we can prove that $\boldsymbol{v} - \boldsymbol{P}\boldsymbol{v}$ is orthogonal to both $\boldsymbol{P}\boldsymbol{v}$ and the column space of $\boldsymbol{X}$, we can prove that matrix $\boldsymbol{P}$ project $\boldsymbol{v}$ onto column space of $\boldsymbol{X}$.
So this problem is equvalent to prove:

$$\begin{aligned}(\boldsymbol{P}\boldsymbol{v})^T(\boldsymbol{v} - \boldsymbol{P}\boldsymbol{v}) &= 0\\ \boldsymbol{X}^T(\boldsymbol{v} - \boldsymbol{P}\boldsymbol{v}) &= 0\end{aligned}$$

Proof:

$$\begin{aligned}(\boldsymbol{P}\boldsymbol{v})^T(\boldsymbol{v} - \boldsymbol{P}\boldsymbol{v}) &= (\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v})^T\boldsymbol{v}\\ &\quad - (\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v})^T(\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v})\\ &= \boldsymbol{v}^T\boldsymbol{X}((\boldsymbol{X}^T\boldsymbol{X})^{-1})^T\boldsymbol{X}^T\boldsymbol{v}\\ &\quad - \boldsymbol{v}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{v}\\ &= \boldsymbol{v}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}\\ &\quad - \boldsymbol{v}^T\boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}\\ &= 0\end{aligned}$$

$$\begin{aligned}\boldsymbol{X}^T(\boldsymbol{v} - \boldsymbol{P}\boldsymbol{v}) &= \boldsymbol{X}^{\mathrm{T}}\boldsymbol{v} - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}^T\boldsymbol{P}\boldsymbol{v}\\ &= \boldsymbol{X}^{\mathrm{T}}\boldsymbol{v} \ - \ \boldsymbol{X}^T\ \boldsymbol{X}\ (\boldsymbol{X}^T\ \boldsymbol{X})^{-1}\ \boldsymbol{X}^T\ \boldsymbol{v}\\ &= \boldsymbol{X}^{\mathrm{T}}\boldsymbol{v} - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{v}\\ &= 0\end{aligned}$$

Thus, $\boldsymbol{P} = \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}$ project $\boldsymbol{v}$ onto column space of $\boldsymbol{X}$.

Then need to prove that $\boldsymbol{\theta}$ correspond to an orthogonal projection of the vector $\boldsymbol{y}$ onto the column space of $\boldsymbol{X}$.

2.2. Then to find the maximum value of l with bringing these constrains into equation using lagrange multiplier

$$
\begin{aligned}
g(l, \lambda, \lambda_{jk}) = \sum_{i=1}^{m} \sum_{j=1}^{d} \log \phi_j(x_j^{(i)} \mid y^{(i)}) + \sum_{i=1}^{m} \log \phi_{y^{(i)}} \\
+\lambda(\sum_{y=1}^{k} \phi_y - 1) + \lambda_{jk}(\sum_{x\in\{0,1\}} \phi_j(x \mid k) - 1)
\end{aligned}
$$

2.3. Questions about SVM

(a) half done

(b) From the KKT condition, here exists:

$$
\begin{aligned}
& \sum_{i=1}^{l} \alpha_i^\star \left[y_i \left(\boldsymbol{w}^{\star \mathrm{T}} \boldsymbol{x}_i + b^\star \right) - 1 \right] = 0 \\
\Rightarrow \quad & \sum_{i=1}^{l} \alpha_i^\star y_i \boldsymbol{w}^{\star \mathrm{T}} \boldsymbol{x}_i + \sum_{i=1}^{l} \alpha_i^\star y_i b^\star = \sum_{i=1}^{l} \alpha_i^\star \\
\Rightarrow \quad & \sum_{i=1}^{l} \alpha_i^\star y_i \boldsymbol{w}^{\star \mathrm{T}} \boldsymbol{x}_i + b^\star \sum_{i=1}^{l} \alpha_i^\star y_i = \sum_{i=1}^{l} \alpha_i^\star \\
\Rightarrow \quad & \sum_{i=1}^{l} \alpha_i^\star y_i \boldsymbol{w}^{\star \mathrm{T}} \boldsymbol{x}_i = \sum_{i=1}^{l} \alpha_i^\star \\
\Rightarrow \quad & \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i^\star \alpha_j^\star y_i y_j \left\langle \boldsymbol{x}_i, \boldsymbol{x}_j \right\rangle = \sum_{i=1}^{l} \alpha_i^\star
\end{aligned}
$$

Then, using the equation above, it has

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

2.4. (a) a

(b) To prove a convex function

$$
f(\boldsymbol{\omega}, b) = \frac{1}{2} \|\boldsymbol{w}\|_2^2 + C \sum_{i=1}^{l} \ell \left(y_i, \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i + b \right)
$$

proof:

$$
\begin{aligned}
&\|\theta \boldsymbol{w}_1 + (1-\theta)\boldsymbol{w}_2\|_2^2 - \theta\|\boldsymbol{w}_1\|_2^2 - (1-\theta)\|\boldsymbol{w}_2\|_2^2 \\
=&\theta^2\|\boldsymbol{w}_1\|_2^2 + 2\theta(1-\theta)\boldsymbol{w}_1^T\boldsymbol{w}_2 + (1-\theta)^2\|\boldsymbol{w}_2\|_2^2 - \theta\|\boldsymbol{w}_1\|_2^2 - (1-\theta)\|\boldsymbol{w}_2\|_2^2 \\
=&2\theta(1-\theta)\boldsymbol{w}_1^T\boldsymbol{w}_2 - \theta(1-\theta)\|\boldsymbol{w}_1\|_2^2 - \theta(1-\theta)\|\boldsymbol{w}_2\|_2^2 \\
\leq&2\theta(1-\theta)\boldsymbol{w}_1^T\boldsymbol{w}_2 - \theta\|\boldsymbol{w}_1\|_2^2 - \theta\|\boldsymbol{w}_2\|_2^2 \\
\leq&-\theta\|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_2^2 \\
\leq&0 \\
\Rightarrow\quad &\|\theta \boldsymbol{w}_1 + (1-\theta)\boldsymbol{w}_2\|_2^2 \leq \theta\|\boldsymbol{w}_1\|_2^2 + (1-\theta)\|\boldsymbol{w}_2\|_2^2
\end{aligned}
$$

So $\|\boldsymbol{w}\|_2^2$ is a convex function.

$$
\begin{aligned}
&\ell\left(y_i, (\theta\boldsymbol{w}_1 + (1-\theta)\boldsymbol{w}_2)^T\boldsymbol{x_i} + \theta b_1 + (1-\theta)b_2\right) \\
=&\max\{1 - y_i\left((\theta\boldsymbol{w}_1 + (1-\theta)\boldsymbol{w}_2)^T\boldsymbol{x_i} + \theta b_1 + (1-\theta)b_2\right), 0\} \\
\leq&\max\{\theta - y_i(\theta\boldsymbol{w}_1^T\boldsymbol{x_i} + \theta b_1) + (1-\theta) - y_i((1-\theta)\boldsymbol{w}_2^T\boldsymbol{x_i} + (1-\theta)b_2), 0\} \\
\leq&\max\{\theta - y_i(\theta\boldsymbol{w}_1^T\boldsymbol{x_i} + \theta b_1)), 0\} + \max\{(1-\theta) - y_i((1-\theta)\boldsymbol{w}_2^T\boldsymbol{x_i} + (1-\theta)b_2), 0\} \\
\leq&\theta\max\{1 - y_i(\boldsymbol{w}_1^T\boldsymbol{x_i} + b_1)), 0\} + (1-\theta)\max\{1 - y_i\boldsymbol{w}_2^T\boldsymbol{x_i} + b_2), 0\} \\
\leq&\theta\ell\left(y_i, \boldsymbol{w}_1^T\boldsymbol{x}_i + b_1\right) + (1-\theta)\ell\left(y_i, \boldsymbol{w}_2^T\boldsymbol{x}_i + b_2\right)
\end{aligned}
$$

So $\ell(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b)$ is a convex function.
The non-negative weighted sum of convex functions is still a convex function. And $C \geq 0$.
Thus the objective function
$f(\boldsymbol{\omega}, b) = \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{l}\ell\left(y_i, \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b\right)$ is convex.