Tsinghua-Berkeley Shenzhen Institute
Learning from Data
Fall 2019

Writing Homework 3

TIAN Chenyu November 14, 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.

3.1. (a) i. Since

$$
\begin{aligned}
&\arg\min_C \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \|\boldsymbol{x}-\boldsymbol{\mu}_j\|^2 \\
\Rightarrow &\arg\min_C \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} (\|\boldsymbol{x}\|^2 + \frac{1}{|C_j|^2}\left\|\sum_{\boldsymbol{x}'\in C_j}\boldsymbol{x}'\right\|^2 - \frac{2}{|C_j|}\boldsymbol{x}^T(\sum_{\boldsymbol{x}'\in C_j}\boldsymbol{x}')) \\
\Rightarrow &\arg\min_C \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \|\boldsymbol{x}\|^2 + \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \frac{1}{|C_j|^2}\left\|\sum_{\boldsymbol{x}'\in C_j}\boldsymbol{x}'\right\|^2 - \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \frac{2}{|C_j|}\boldsymbol{x}^T(\sum_{\boldsymbol{x}'\in C_j}\boldsymbol{x}') \\
\Rightarrow &\arg\min_C \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \|\boldsymbol{x}\|^2 + \sum_{j=1}^{k} \frac{1}{|C_j|}\left\|\sum_{\boldsymbol{x}'\in C_j}\boldsymbol{x}'\right\|^2 - \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \frac{2}{|C_j|}\boldsymbol{x}^T(\sum_{\boldsymbol{x}'\in C_j}\boldsymbol{x}') \\
\Rightarrow &\arg\min_C \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \|\boldsymbol{x}\|^2 - \sum_{j=1}^{k} \frac{1}{|C_j|}\left\|\sum_{\boldsymbol{x}'\in C_j}\boldsymbol{x}'\right\|^2 \\
\Rightarrow &\arg\min_C \sum_{j=1}^{k} (\sum_{\boldsymbol{x}\in C_j} \|\boldsymbol{x}\|^2 - \frac{1}{|C_j|}\sum_{\boldsymbol{x},\boldsymbol{x}'\in C_j}\boldsymbol{x}^T\boldsymbol{x}')
\end{aligned}
$$

ii. Based on

$$
\begin{aligned}
&\arg\min_C \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \|\boldsymbol{x}-\boldsymbol{\mu}_j\|^2 \\
\Rightarrow &\arg\min_C \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} (\|\boldsymbol{x}\|^2 + \frac{1}{|C_j|^2}\left\|\sum_{\boldsymbol{x}'\in C_j} \boldsymbol{x}\right\|^2 - \frac{2}{|C_j|}\boldsymbol{x}^T(\sum_{\boldsymbol{x}'\in C_j} \boldsymbol{x})) \\
\Rightarrow &\arg\min_C \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \|\boldsymbol{x}\|^2 + \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \frac{1}{|C_j|^2}\left\|\sum_{\boldsymbol{x}'\in C_j} \boldsymbol{x}\right\|^2 - \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \frac{2}{|C_j|}\boldsymbol{x}^T(\sum_{\boldsymbol{x}'\in C_j} \boldsymbol{x}) \\
\Rightarrow &\arg\min_C \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \|\boldsymbol{x}\|^2 + \sum_{j=1}^{k} \frac{1}{|C_j|}\left\|\sum_{\boldsymbol{x}'\in C_j} \boldsymbol{x}\right\|^2 - \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \frac{2}{|C_j|}\boldsymbol{x}^T(\sum_{\boldsymbol{x}'\in C_j} \boldsymbol{x}) \\
\Rightarrow &\arg\min_C \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \|\boldsymbol{x}\|^2 - \sum_{j=1}^{k} \frac{1}{|C_j|}\left\|\sum_{\boldsymbol{x}'\in C_j} \boldsymbol{x}\right\|^2 \\
\Rightarrow &\arg\max_C \sum_{j=1}^{k} \frac{1}{|C_j|}\left\|\sum_{\boldsymbol{x}'\in C_j} \boldsymbol{x}\right\|^2 - \sum_{j=1}^{k} \sum_{\boldsymbol{x}\in C_j} \|\boldsymbol{x}\|^2 \\
\Rightarrow &\arg\max_C \sum_{j=1}^{k} \frac{1}{|C_j|} \sum_{\boldsymbol{x},\boldsymbol{x}'\in C_j} \boldsymbol{x}^T\boldsymbol{x}'
\end{aligned}
$$

(b) Based on

i. Since the

ii. Based on

3.2. (a) Since the

i. Since the

ii. Based on

(b) Based on

$$
\boldsymbol{\Sigma_1} = \frac{1}{\sum_{i=1}^{m} \mathbb{1}(y^{(i)}=1)} \sum_{i=1}^{m} \mathbb{1}(y^{(i)}=1)\left(\boldsymbol{x}^{(i)}-\boldsymbol{\mu_1}\right)\left(\boldsymbol{x}^{(i)}-\boldsymbol{\mu_1}\right)^{\mathrm{T}}
$$

3.3. (a) Since the

(b) baskfdjl

(c) Based on

3.4.

3.5. (a) The original problem is

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

For the optimal solution,
if $y_i\left(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i+b\right) \geq 1$, because we want to minimize $\sum_{i=1}^l \xi_i$, ξ_i must be 0, which equals to $\ell\left(y_i, \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i+b\right)$;
if $y_i\left(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i+b\right)<1$, because of the constrains, ξ_i must be $1-y_i\left(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i+b\right)$, which equals to $\ell\left(y_i, \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i+b\right)$.
This means if find the solution of the original problem, the solution of (3) in file *wa2* is found.

(b) To prove a convex function

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

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.