

Writing Homework 2

TIAN Chenyu

October 29, 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 $\mathbf{P} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$, for a given vector \mathbf{v}

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

If we can prove that $\mathbf{v} - \mathbf{P}\mathbf{v}$ is orthogonal to both $\mathbf{P}\mathbf{v}$ and the column space of \mathbf{X} , we can prove that matrix \mathbf{P} project \mathbf{v} onto column space of \mathbf{X} .

So this problem is equivalent to prove:

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

Proof:

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

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

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

So, $\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\theta} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \mathbf{P}\mathbf{y}$ correspond to an orthogonal projection of the vector \mathbf{y} onto the column space of \mathbf{X} .

2.2.

$$p(\mathbf{x}|y=0) = \frac{1}{(2\pi)^{n/2}|\Sigma_0|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_0)^T \Sigma_0^{-1}(\mathbf{x} - \mu_0)\right)$$

$$p(\mathbf{x}|y=1) = \frac{1}{(2\pi)^{n/2}|\Sigma_1|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_1)^T \Sigma_1^{-1}(\mathbf{x} - \mu_1)\right)$$

The log likelihood function of QDA is

$$l(\phi, \mu_0, \mu_1, \Sigma_0, \Sigma_1) = \log \prod_{i=1}^m p(\mathbf{x}^{(i)}, y^{(i)}; \phi, \mu_0, \mu_1, \Sigma_0, \Sigma_1)$$

$$= \log \prod_{i=1}^m p(\mathbf{x}^{(i)}|y^{(i)}; \mu_0, \mu_1, \Sigma_0, \Sigma_1) \phi_{y^{(i)}}$$

For Σ_0 , we have

$$\begin{aligned} \frac{\partial l(\phi, \mu_0, \mu_1, \Sigma_0, \Sigma_1)}{\partial \Sigma_0} &= -\frac{\sum_{i=1}^m \mathbb{1}(y^{(i)}=0)}{2} \frac{\partial}{\partial \Sigma_0} \log |\Sigma_0| \\ &\quad - \frac{1}{2} \frac{\partial}{\partial \Sigma_0} \sum_{i=1}^m \mathbb{1}(y^{(i)}=0) (\mathbf{x}^{(i)} - \mu_0)^T \Sigma_0^{-1} (\mathbf{x}^{(i)} - \mu_0) \\ &= -\frac{\sum_{i=1}^m \mathbb{1}(y^{(i)}=0)}{2} \Sigma_0^{-1} \\ &\quad + \frac{1}{2} \Sigma_0^{-1} \left[\sum_{i=1}^m \mathbb{1}(y^{(i)}=0) (\mathbf{x}^{(i)} - \mu_0) (\mathbf{x}^{(i)} - \mu_0)^T \right] \Sigma_0^{-1} \\ &= O \end{aligned}$$

which yields that

$$\Sigma_0 = \frac{1}{\sum_{i=1}^m \mathbb{1}(y^{(i)}=0)} \sum_{i=1}^m \mathbb{1}(y^{(i)}=0) (\mathbf{x}^{(i)} - \mu_0) (\mathbf{x}^{(i)} - \mu_0)^T$$

With same derivation

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

2.3. (a) Since the data is separable, there exist support vectors which

$$y_i (\mathbf{w}^T \mathbf{x}_i + b) = 1.$$

When $y_i = 1$, has constrain $\mathbf{w}^T \mathbf{x}_i + b \geq 1$, and $\min_{i:y_i=1} \mathbf{w}^*{}^T \mathbf{x}_i + b^* = 1$;

When $y_i = -1$, has constrain $\mathbf{w}^T \mathbf{x}_i + b \leq -1$, and

$$\max_{i:y_i=-1} \mathbf{w}^*{}^T \mathbf{x}_i + b^* = -1;$$

Therefore,

$$\begin{aligned} &\max_{i:y_i=-1} \mathbf{w}^*{}^T \mathbf{x}_i + b^* + \min_{i:y_i=1} \mathbf{w}^*{}^T \mathbf{x}_i + b^* = 0 \\ \Rightarrow &b^* = -\frac{1}{2} \left(\max_{i:y_i=-1} \mathbf{w}^*{}^T \mathbf{x}_i + \min_{i:y_i=1} \mathbf{w}^*{}^T \mathbf{x}_i \right) \end{aligned}$$

(b) Based on the KKT condition, here exists:

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

Then, using the equation above, it has

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

2.4. (a) The original problem is

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

For the optimal solution,

if $y_i (\mathbf{w}^{\text{T}} \mathbf{x}_i + b) \geq 1$, because we want to minimize $\sum_{i=1}^l \xi_i$, ξ_i must be 0, which equals to $\ell(y_i, \mathbf{w}^{\text{T}} \mathbf{x}_i + b)$;

if $y_i (\mathbf{w}^{\text{T}} \mathbf{x}_i + b) < 1$, because of the constrains, ξ_i must be $1 - y_i (\mathbf{w}^{\text{T}} \mathbf{x}_i + b)$, which equals to $\ell(y_i, \mathbf{w}^{\text{T}} \mathbf{x}_i + b)$.

This means if find the solution of the original problem, the solution of (3) in file *wa2* is found. Thus, the problem is equivalent to

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

(b) To prove a convex function

$$f(\mathbf{w}, b) = \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^l \ell(y_i, \mathbf{w}^{\text{T}} \mathbf{x}_i + b)$$

proof:

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

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

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

So $\ell(\mathbf{w}^T \mathbf{x}_i + b)$ 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

$$f(\mathbf{w}, b) = \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^l \ell(y_i, \mathbf{w}^T \mathbf{x}_i + b) \text{ is convex.}$$