Duality and kernel algorithm

Chanwoo Lee, August 19, 2020

I derive dual problem based on concatenated map first and think bilinear map later.

1 Previous duality derivation in concatenated mapping

Primal problem:

$$\min_{\boldsymbol{P},\boldsymbol{C},\boldsymbol{\xi}} \frac{1}{2}\|\boldsymbol{C}\boldsymbol{P}^T\|_F^2 + c\sum_{i=1}^n \xi_i \tag{1}$$
$$\text{subject to } y_i\langle \boldsymbol{C}\boldsymbol{P}^T, \Phi(\boldsymbol{X}_i)\rangle \leq 1-\xi_i \text{ and } \xi_i \geq 0, i = 1,\ldots,n.$$

We can have equivalent problem introducing dummy variables $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_n), \boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_n)$.

Primal problem:

$$\min_{\boldsymbol{P},\boldsymbol{C},\boldsymbol{\xi}} \max_{\boldsymbol{\alpha},\boldsymbol{\mu}} \frac{1}{2}\|\boldsymbol{C}\boldsymbol{P}^T\|_F^2 + c\sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i\left[(y_i\langle \boldsymbol{C}\boldsymbol{P}^T, \Phi(\boldsymbol{X}_i)\rangle - (1-\xi_i)\right] - \sum_{i=1}^n \mu_i\xi_i. \tag{2}$$

(1) and (2) are equivalent because "$\max_{\boldsymbol{\alpha},\boldsymbol{\mu}}$" part force $\boldsymbol{P},\boldsymbol{C}$ satisfy the constraint in (1). From (2), we can have dual problem.

Dual problem:

$$\max_{\boldsymbol{\alpha},\boldsymbol{\mu}} \min_{\boldsymbol{P},\boldsymbol{C},\boldsymbol{\xi}} \frac{1}{2}\|\boldsymbol{C}\boldsymbol{P}^T\|_F^2 + c\sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i\left[(y_i\langle \boldsymbol{C}\boldsymbol{P}^T, \Phi(\boldsymbol{X}_i)\rangle - (1-\xi_i)\right] - \sum_{i=1}^n \mu_i\xi_i. \tag{3}$$

From (3), we focus on "$\min_{\boldsymbol{P},\boldsymbol{C},\boldsymbol{\xi}}$" part. We use alternating optimization fixing $\boldsymbol{P}$. Then, we derive first order condition on $(\boldsymbol{P},\boldsymbol{C},\boldsymbol{\xi})$ fixing $\boldsymbol{P}$, and replace $(\boldsymbol{C},\boldsymbol{\xi})$ in terms of $(\boldsymbol{\alpha},\boldsymbol{\mu})$. We have the following dual problem when we fix $\boldsymbol{P}$.

Dual problem:

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j \langle \Phi(\boldsymbol{X}_i)\boldsymbol{P}, \Phi(\boldsymbol{X}_i)\boldsymbol{P}\rangle, \tag{4}$$
$$\text{subject to } \sum_i y_i\alpha_i = 0, \text{ and } 0\leq \alpha_i \leq c, i = 1,\ldots,n.$$

Notice that the constraint in (4) is from the first order condition on $(\boldsymbol{C},\boldsymbol{\xi})$. When $\boldsymbol{P}$ fixed, strong duality holds such that

$$\min_{\boldsymbol{C},\boldsymbol{\xi}} \max_{\boldsymbol{\alpha},\boldsymbol{\mu}} L(\boldsymbol{C},\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\mu}) = \max_{\boldsymbol{\alpha},\boldsymbol{\mu}} \min_{\boldsymbol{C},\boldsymbol{\xi}} L(\boldsymbol{C},\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\mu})$$

where $L(\boldsymbol{C},\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\mu}) = \|\boldsymbol{C}\boldsymbol{P}^T\|_F^2 + c\sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \left[(y_i\langle \boldsymbol{C}\boldsymbol{P}^T, \Phi(\boldsymbol{X}_i)\rangle - (1-\xi_i)\right] - \sum_{i=1}^n \mu_i\xi_i$,

and $\boldsymbol{P}$ is assumed to be fixed.

By the same way, we derived the dual problem when $\boldsymbol{C}$ is fixed.

2 Issue in Section 4

We can not derive the dual problem (9) in the previous note (1).

$$\max_{\boldsymbol{P}}\min_{\boldsymbol{\alpha}} -\sum_{i=1}^n \alpha_i + \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j \langle \Phi(\boldsymbol{X}_i)\boldsymbol{P}, \Phi(\boldsymbol{X}_i)\boldsymbol{P}\rangle,$$
$$\text{subject to } \sum_i y_i\alpha_i = 0, \text{ and } 0 \le \alpha_i \le c, i = 1,\dots,n.$$

The dual problem of (1) is (3)

$$\max_{\boldsymbol{\alpha},\boldsymbol{\mu}}\min_{\boldsymbol{P},\boldsymbol{C},\boldsymbol{\xi}} \frac{1}{2}\|\boldsymbol{C}\boldsymbol{P}^T\|_F^2 + c\sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \left[(y_i\langle \boldsymbol{C}\boldsymbol{P}^T, \Phi(\boldsymbol{X}_i)\rangle - (1-\xi_i)\right] - \sum_{i=1}^n \mu_i\xi_i.$$

First, let "$\min_{\boldsymbol{P},\boldsymbol{C},\boldsymbol{\xi}}$" reduced to "$\min_{\boldsymbol{P}}$". If we do not want to fix $\boldsymbol{P}$, then by the first order condition $(\boldsymbol{P},\boldsymbol{C},\boldsymbol{\xi})$ must satisfy

$$\begin{aligned}
&\boldsymbol{C} = \sum_{i=1}^n \alpha_i y_i \Phi(\boldsymbol{X}_i)\boldsymbol{P},\\
&\boldsymbol{P} = \sum_{i=1}^n \alpha_i y_i (\boldsymbol{C}^T\boldsymbol{C})^{-1}\boldsymbol{C}^T\Phi(\boldsymbol{X}_i),\\
&c\mathbb{1} + \boldsymbol{\alpha} - \boldsymbol{\mu} = 0.
\end{aligned}$$

If we plug the first and the second constraint into (3), we have

$$\min_{\boldsymbol{\alpha}}\max_{\boldsymbol{P}} -\sum_{i=1}^n \alpha_i + \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j \langle \Phi(\boldsymbol{X}_i)\boldsymbol{P}, \Phi(\boldsymbol{X}_i)\boldsymbol{P}\rangle, \tag{5}$$
$$\text{subject to } \sum_i y_i\alpha_i = 0, \text{ and } 0 \le \alpha_i \le c, i = 1,\dots,n.$$

However, $\boldsymbol{P}$ also should satisfy the first order condition. Therefore, we cannot say (5) is equivalent to (3). Furthermore, when we do not use alternating algorithm where one of $\boldsymbol{P}$ and $\boldsymbol{C}$ is fixed, there is no guarantee that the strong duality hold.

Numerically, The table is the values of (5) on each iteration which shows non-monotonic pattern.

3 Questions about bilinear mapping

1. In primal problem(Equation (1)) in the previous note, penalty part on function complexity is $\|\boldsymbol{C}\|_F^2$ not $\|\boldsymbol{P}_r\boldsymbol{C}\boldsymbol{P}_c^T\|_F^2$. I am not sure that the case where the penalty has $\|\boldsymbol{C}\|_F^2$ can be reduced to linear kernel case with low rank coefficient.

Iteration	1	2	3	4	5	6	7
Loss	44.18116	11.90551	49.82844	22.60086	41.06655	34.81242	62.36754

Table 1: Numerical result of loss function value on each iteration. Loss is calculated based on the loss in Equation (5)

2. In $\|\boldsymbol{C}\|_F^2$ case, the dual solution is

$$\max_{\boldsymbol{\alpha},\boldsymbol{\mu}} \min_{\boldsymbol{P}_r,\boldsymbol{P}_c,\boldsymbol{C},\boldsymbol{\xi}} \frac{1}{2}\|\boldsymbol{C}\|_F^2 + c\sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \left[(y_i\langle \boldsymbol{C}, \boldsymbol{P}_r\Phi(\boldsymbol{X}_i)\boldsymbol{P}_c^T\rangle - (1-\xi_i)\right] - \sum_{i=1}^n \mu_i\xi_i.$$

By the similar argument in previous section, we can derive dual problem when $\boldsymbol{P}_r, \boldsymbol{P}_c$ fixed

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j y_i y_j \langle \boldsymbol{P}_r^T\Phi(\boldsymbol{X}_i)\boldsymbol{P}_c, \boldsymbol{P}_r^T\Phi(\boldsymbol{X}_j)\boldsymbol{P}_c\rangle,$$
$$\text{subject to } \sum_{i=1}^n y_i\alpha_i \text{ and } y_i\alpha_i = 0, \text{ and } 0 \le \alpha_i \le c, i = 1,\ldots,n.$$

However, $\boldsymbol{P}_r$ and $\boldsymbol{P}_c$ are not fixed, I cannot derive

$$\begin{aligned}
&\min_{\boldsymbol{P}_r,\boldsymbol{P}_c,\boldsymbol{C},\boldsymbol{\xi}} \frac{1}{2}\|\boldsymbol{C}\|_F^2 + c\sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \left[(y_i\langle \boldsymbol{C}, \boldsymbol{P}_r\Phi(\boldsymbol{X}_i)\boldsymbol{P}_c^T\rangle - (1-\xi_i)\right] - \sum_{i=1}^n \mu_i\xi_i \\
&= \min_{\boldsymbol{P}_r,\boldsymbol{P}_c} \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j y_i y_j \langle \boldsymbol{P}_r^T\Phi(\boldsymbol{X}_i)\boldsymbol{P}_c, \boldsymbol{P}_r^T\Phi(\boldsymbol{X}_j)\boldsymbol{P}_c\rangle.
\end{aligned}$$