Multiple Instance Learning on Structured Data – Supplemental Materials

In this part, we will prove Theorem 1 and Theorem 2 in the paper.

Theorem 1: Any solution $\mathbf{w}^*$ of problem (4) is also a solution to problem (5) (and vice versa) with $\xi^* = \frac{1}{n}\sum_{i=1}^{n}\xi_i^*$ and $\zeta^* = \frac{1}{|E|}\sum_{(p,q)\in E}\zeta_{(p,q)}^*$.

Proof: Adapting the proof from Joachims (2006), we will prove that for every $\mathbf{w}$, the smallest feasible ξ^* and ζ^* are related by $\xi^* = \frac{1}{n}\sum_{i=1}^{n}\xi_i^*$ and $\zeta^* = \frac{1}{|E|}\sum_{(p,q)\in E}\zeta_{(p,q)}^*$. For any given $\mathbf{w}$, the optimal ξ_i^* in problem (4) can be achieved individually as $\xi_i^* = \max\{0, 1 - Y_i\mathbf{w}^T\mathbf{B}_{iu_i^*}\}$ and the optimal ξ^* in problem (5) is achieved by $\xi^* = \max_{\mathbf{c}_k\in\{0,1\}}\{0, \frac{1}{n}\sum_{k=1}^{n}\mathbf{c}_k - \frac{1}{n}\mathbf{w}^T\sum_{k=1}^{n}\mathbf{c}_kY_k\mathbf{B}_{ku_k^*}\} = \frac{1}{n}\sum_{i=1}^{n}\max\{0, 1 - Y_i\mathbf{w}^T\mathbf{B}_{iu_i^*}\} = \frac{1}{n}\sum_{i=1}^{n}\xi_i^*$. Similarly,
$\zeta_{(p,q)}^* = \max\left\{0, \max_{k\in\{1,\ldots,n_p\}}\{\frac{\mathbf{w}^T\mathbf{B}_{pk}}{\sqrt{d(p)}} - \frac{\mathbf{w}^T\mathbf{B}_{qu_q^*}}{\sqrt{d(q)}}\}, \max_{k\in\{n_p+1,\ldots,n_p+n_q\}}\{\frac{\mathbf{w}^T\mathbf{B}_{q(k-n_p)}}{\sqrt{d(q)}} - \frac{\mathbf{w}^T\mathbf{B}_{pu_p^*}}{\sqrt{d(p)}}\}\right\}$
and

$$\begin{aligned}
\zeta^* &= \max\left\{0, \frac{1}{|E|}\mathbf{w}^T\sum_{j=1}^{|E|}\left(\sum_{k=1}^{n_p}\tau_{jk}(\frac{\mathbf{B}_{pk}}{\sqrt{d(p)}} - \frac{\mathbf{B}_{qu_q^*}}{\sqrt{d(q)}}) + \sum_{k=1}^{n_q}\tau_{jk}(\frac{\mathbf{B}_{qk}}{\sqrt{d(q)}} - \frac{\mathbf{B}_{pu_p^*}}{\sqrt{d(p)}})\right)\right\} \\
&= \frac{1}{|E|}\sum_{j=1}^{|E|}\max\left\{0, \max_{k\in\{1,\ldots,n_p\}}\{\frac{\mathbf{w}^T\mathbf{B}_{pk}}{\sqrt{d(p)}} - \frac{\mathbf{w}^T\mathbf{B}_{qu_q^*}}{\sqrt{d(q)}}\}, \max_{k\in\{n_p+1,\ldots,n_p+n_q\}}\{\frac{\mathbf{w}^T\mathbf{B}_{q(k-n_p)}}{\sqrt{d(q)}} - \frac{\mathbf{w}^T\mathbf{B}_{pu_p^*}}{\sqrt{d(p)}}\}\right\} \\
&= \frac{1}{|E|}\sum_{i=1}^{|E|}\zeta_{(p,q)}^*
\end{aligned}$$

where, $e_j = (p,q)$. So, for any $\mathbf{w}$, the objective function values for problem (4) and (5) are the same given the optimal ξ, ξ_i, ζ and $\zeta_{(p,q)}$. Therefore, these two problems are exactly equivalent.

□

Theorem 2: For each cutting plane iteration described in Table 1, the objective function will be increased by at least $\kappa = \min\{\frac{C\epsilon_1}{2}, \frac{\epsilon_1^2}{8R^2}, \frac{\mu\epsilon_2}{2}, \frac{\epsilon_2^2}{16R^2}, \frac{(\epsilon_1+\epsilon_2)^2}{(24+16\sqrt{2})R^2}\}$, and $R^2 = \max_{i,j} \mathbf{B}_{ij}^2$.

Proof: The proof is based on the dual form of the original problem. So, we first wrote down the dual form and then proved the theorem based on this dual form.

For the s-th Cutting Plane iteration, the optimization problem can be written as:

$$\min_{\mathbf{w},\xi\geq 0,\zeta\geq 0} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\xi + \mu\zeta \tag{1}$$

$$s.t. \quad \forall \mathbf{c}_i \in \Omega_1^{t_s}, \quad \frac{1}{n}\mathbf{w}^T\sum_{k=1}^{n}\mathbf{c}_{ik}Y_k\mathbf{B}_{ku_k^{(t)}} \geq \frac{1}{n}\sum_{k=1}^{n}\mathbf{c}_{ik} - \xi$$

$$\forall \tau_i \in \Omega_2^{t_s}, \frac{1}{|E|}\mathbf{w}^T\sum_{j=1}^{|E|}\left(\sum_{k=1}^{n_p}\tau_{ijk}(\frac{\mathbf{B}_{pk}}{\sqrt{d(p)}} - \frac{\mathbf{B}_{qu_q^{(t)}}}{\sqrt{d(q)}}) + \sum_{k=1}^{n_q}\tau_{ij(k+n_p)}(\frac{\mathbf{B}_{qk}}{\sqrt{d(q)}} - \frac{\mathbf{B}_{pu_p^{(t)}}}{\sqrt{d(p)}})\right) \leq \zeta$$

Please note that $e_j = (p,q)$, for the convenience of analysis, let's assume that $\mathbf{x}_i^{t_s} = \frac{1}{n}\sum_{k=1}^{n} c_{ik}Y_k\mathbf{B}_{ku_k^{(t)}}, \forall c_i \in \Omega_1^{t_s}$, and $\mathbf{x}_{i+|\Omega_1^{t_s}|}^{t_s} = -\frac{1}{|E|}\sum_{j=1}^{|E|}\left(\sum_{k=1}^{n_p}\tau_{ijk}(\frac{\mathbf{B}_{pk}}{\sqrt{d(p)}} - \frac{\mathbf{B}_{qu_q^{(t)}}}{\sqrt{d(q)}}) + \sum_{k=1}^{n_q}\tau_{ij(k+n_p)}(\frac{\mathbf{B}_{qk}}{\sqrt{d(q)}} - \frac{\mathbf{B}_{pu_p^{(t)}}}{\sqrt{d(p)}})\right), \quad \forall \tau_i \in \Omega_2^{t_s}$.

The transformed formulation is:

$$\min_{\mathbf{w},\xi\geq 0,\zeta\geq 0} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\xi + \mu\zeta \tag{2}$$

$$s.t. \quad \forall i \in \{1,2,\ldots,|\Omega_1^{t_s}|\}, \quad \mathbf{w}^T\mathbf{x}_i^{t_s} \geq \frac{1}{n}\mathbf{c}_i^T\mathbf{1} - \xi$$

$$\forall i \in \{|\Omega_1^{t_s}|+1, |\Omega_1^{t_s}|+2, \ldots, |\Omega_1^{t_s}|+|\Omega_2^{t_s}|\}, \quad \mathbf{w}^T\mathbf{x}_i^{t_s} \geq -\zeta$$

The dual form of problem is:

$$\max_{\alpha \geq 0} \quad -\frac{1}{2}\alpha^T \mathbf{K}\alpha + \frac{1}{n}\sum_{i=1}^{|\Omega_1^{t_s}|} \alpha_i \mathbf{c}_i^T \mathbf{1} \tag{3}$$
$$s.t. \quad \sum_{i=1}^{|\Omega_1^{t_s}|} \alpha_i = C$$
$$\sum_{i=|\Omega_1^{t_s}|+1}^{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|} \alpha_i = \mu$$

where $\mathbf{K} = \mathbf{X}^T\mathbf{X}$ and $\mathbf{X} = [\mathbf{x}_1^{t_s}, \ldots, \mathbf{x}_{|\Omega_1^{t_s}|}^{t_s}, \ldots, \mathbf{x}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|}]$

For each iteration, there are three possibilities:

1.$H_1^{t_s}$ is false. $H_2^{t_s}$ is true. $\mathbf{c}^{t_s}$ is added to $\Omega_1^{t_s}$

Let $\mathbf{c}^{t_{s+1}}$ be the newly added constraint. Let $\widehat{\alpha}^{t_s}$ be the optimal solution of the dual problem before this constraint is added.

Suppose $\widetilde{\alpha}^{t_s} = [\widetilde{\alpha}_1^{t_s}, \widetilde{\alpha}_2^{t_s}, \ldots, \widetilde{\alpha}_{|\Omega_1^{t_s}|}^{t_s}, 0, \widetilde{\alpha}_{|\Omega_1^{t_s}|+1}^{t_s}, \widetilde{\alpha}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|}^{t_s}]$. To lower bound the progress made by each cutting plane iteration, we consider the increase in a specific direction of the dual function by line search, which can be described as: $\max_{0 \leq \beta \leq C} D^{t_{s+1}}(\widetilde{\alpha}^{t_s} + \beta\eta) - D^{t_{s+1}}(\widetilde{\alpha}^{t_s})$, where the direction η is specified as (with this direction, and $0 \leq \beta \leq C$, the solution will be guaranteed to lie in the feasible region):

$$\eta_j = \begin{cases} -\frac{1}{C}\widetilde{\alpha}_j^{t_s}, & j = 1, \ldots, |\Omega_1^{t_s}| \\ 1, & j = |\Omega_1^{t_s}| + 1 \\ 0, & j = |\Omega_1^{t_s}| + 2, \ldots, |\Omega_1^{t_s}| + |\Omega_2^{t_s}| + 1 \end{cases} \tag{4}$$

In our formulation, we have: $\frac{\partial D^{t_{s+1}}(\widetilde{\alpha}^{t_s})}{\partial \widetilde{\alpha}_j^{t_s}} = \frac{1}{n}(\mathbf{c}^{t_j})^T\mathbf{1} - (\widetilde{\mathbf{w}}^{t_s})^T\mathbf{x}_j^{t_{s+1}} = \xi$, if $\widetilde{\alpha}_j^{t_s} \neq 0, j = 1, \ldots, |\Omega_1^{t_s}|$, and $\frac{\partial D^{t_{s+1}}(\widetilde{\alpha}^{t_s})}{\partial \widetilde{\alpha}_{|\Omega_1^{t_s}|+1}^{t_s}} = \frac{1}{n}(\mathbf{c}^{t_j})^T\mathbf{1} - (\widetilde{\mathbf{w}}^{t_s})^T\mathbf{x}_j^{t_{s+1}} \geq \xi + \epsilon_1$. Then, we can get:

$$\nabla D^{t_{s+1}}(\widetilde{\alpha}^{t_s})^T\eta \geq \xi + \epsilon_1 - \sum_{j=1}^{|\Omega_1^{t_s}|} \frac{\widetilde{\alpha}_j^{t_s}}{C}\xi \geq \epsilon_1 \geq 0$$

$$\eta^T\mathbf{K}\eta =(\mathbf{x}_{|\Omega|+1}^{t_{s+1}}) - \frac{2}{C}\sum_{i=1}^{|\Omega_1^{t_s}|}\tilde{\alpha}_i^{t_s}K(\mathbf{x}_i^{t_{s+1}}, \mathbf{x}_{s+1}^{t_{s+1}}) + \frac{1}{C^2}\sum_{i=1}^{|\Omega_1^{t_s}|}\sum_{j=1}^{|\Omega_1^{t_s}|}\tilde{\alpha}_i^{t_s}\tilde{\alpha}_j^{t_s}K(\mathbf{x}_i^{t_{s+1}}, \mathbf{x}_j^{t_{s+1}})$$

$$\leq R^2 + \frac{2}{C}CR^2 + \frac{1}{C^2}C^2R^2 = 4R^2 \tag{5}$$

where $R^2 = \max_{i,j\in\mathbf{B}_i}\mathbf{B}_{ij}^2$. By utilizing Corollary (13) proved in Tsochantaridis et al. (2006), the minimum increase is:

$$\max_{0\leq\beta\leq C} D^{t_{s+1}}(\tilde{\alpha}^{t_s} + \beta\eta) - D^{t_{s+1}}(\tilde{\alpha}^{t_s}) \geq \min\{\frac{C\epsilon_1}{2}, \frac{\epsilon_1^2}{8R^2}\} \tag{6}$$

2. $H_1^{t_s}$ is true. $H_2^{t_s}$ is false. $\Omega_2^{t_s}$ is updated by appending τ^{t_s}.

Let $\tau^{t_{s+1}}$ be the newly added constraint. Let $\widehat{\alpha}^{t_s}$ be the optimal solution of the dual problem before this constraint is added.

Suppose $\widetilde{\alpha}^{t_s} = [\widetilde{\alpha}_1^{t_s}, \widetilde{\alpha}_2^{t_s}, \dots, \widetilde{\alpha}_{|\Omega_1^{t_s}|}^{t_s}, \widetilde{\alpha}_{|\Omega_1^{t_s}|+1}^{t_s}, \widetilde{\alpha}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|}^{t_s}, 0]$. To lower bound the progress made by each cutting plane iteration, we consider the increase in a specific direction of the dual function by line search, which can be described as: $\max_{0\leq\beta\leq\mu} D^{t_{s+1}}(\widetilde{\alpha}^{t_s} + \beta\eta) - D^{t_{s+1}}(\widetilde{\alpha}^{t_s})$, where the direction η is specified as:

$$\eta_j = \begin{cases} 0, & j = 1, \dots, |\Omega_1^{t_s}| \\ -\frac{1}{\mu}\widetilde{\alpha}_j^{t_s}, & j = |\Omega_1^{t_s}| + 1, \dots, |\Omega_1^{t_s}| + |\Omega_2^{t_s}| \\ 1, & j = |\Omega_1^{t_s}| + |\Omega_2^{t_s}| + 1 \end{cases} \tag{7}$$

In our formulation, we have: $\frac{\partial D^{t_{s+1}}(\widetilde{\alpha}^{t_s})}{\partial\widetilde{\alpha}_j^{t_s}} = -(\widetilde{\mathbf{w}}^{t_s})^T\mathbf{x}_j^{t_{s+1}} = \zeta$, if $\widetilde{\alpha}_j^{t_s} \neq 0$, $j = |\Omega_1^{t_s}| + 1, \dots, |\Omega_1^{t_s}| + |\Omega_2^{t_s}|$, and $\frac{\partial D^{t_{s+1}}(\widetilde{\alpha}^{t_s})}{\partial\widetilde{\alpha}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+1}^{t_s}} = -(\widetilde{\mathbf{w}}^{t_s})^T\mathbf{x}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+1}^{t_{s+1}} \geq (\zeta + \epsilon_2)$. Then, we can get:

$$\nabla D^{t_{s+1}}(\widetilde{\alpha}^{t_s})^T\eta \geq \zeta + \epsilon_2 - \sum_{j=|\Omega_1^{t_s}|+1}^{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|}\frac{\widetilde{\alpha}_j^{t_s}}{\mu}\zeta \geq \epsilon_2 \geq 0$$

$$
\begin{aligned}
\eta^T \mathbf{K} \eta =& (\mathbf{x}^{t_{s+1}}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+1})^2 - \frac{2}{\mu} \sum_{i=|\Omega_1^{t_s}|+1}^{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|} \tilde{\alpha}_i^{t_s} K(\mathbf{x}_i^{t_{s+1}}, \mathbf{x}^{t_{s+1}}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+1}) \\
& + \frac{1}{\mu^2} \sum_{i=|\Omega_1^{t_s}|+1}^{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|} \sum_{j=|\Omega_1^{t_s}|+1}^{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|} \tilde{\alpha}_i^{t_s} \tilde{\alpha}_j^{t_s} K(\mathbf{x}_i^{t_{s+1}}, \mathbf{x}_j^{t_{s+1}}) \\
\le & Q^2 + \frac{2}{\mu} \mu Q^2 + \frac{1}{\mu^2} \mu^2 Q^2 = 4Q^2 = 8R^2
\end{aligned} \tag{8}
$$

By utilizing Corollary 13 proved in Tsochantaridis et al. (2006), the minimum increase is:

$$
\max_{0 \le \beta \le \mu} D^{t_{s+1}}(\tilde{\alpha}^{t_s} + \beta \eta) - D^{t_{s+1}}(\tilde{\alpha}^{t_s}) \ge \min\{\frac{\mu \epsilon_2}{2}, \frac{\epsilon_2^2}{16R^2}\} \tag{9}
$$

3. Both $H_1^{t_s}$ and $H_2^{t_s}$ are false. The most violated constraints are added to both $\Omega_1^{t_s}$ and $\Omega_2^{t_s}$.

Let $\mathbf{c}^{t_{s+1}}$ and $\tau^{t_{s+1}}$ be the newly added constraint. Let $\widehat{\alpha}^{t_s}$ be the optimal solution of the dual problem before this constraint is added.

Suppose $\widetilde{\alpha}^{t_s} = [\widetilde{\alpha}_1^{t_s}, \widetilde{\alpha}_2^{t_s}, \ldots, \widetilde{\alpha}^{t_s}_{|\Omega_1^{t_s}|}, 0, \widetilde{\alpha}^{t_s}_{|\Omega_1^{t_s}|+1}, \widetilde{\alpha}^{t_s}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|}, 0]$. To lower bound the progress made by each cutting plane iteration, we consider the increase in a specific direction of the dual function by line search, which can be described as: $\max_{0 \le \beta \le \min\{C, \mu\}} D^{t_{s+1}}(\widetilde{\alpha}^{t_s} + \beta \eta) - D^{t_{s+1}}(\widetilde{\alpha}^{t_s})$, where the direction η is specified as:

$$
\eta_j = \begin{cases} -\frac{1}{C} \widetilde{\alpha}_j^{t_s}, & j = 1, \ldots, |\Omega_1^{t_s}| \\ 1, & j = |\Omega_1^{t_s}| + 1 \\ -\frac{1}{\mu} \widetilde{\alpha}_j^{t_s}, & j = |\Omega_1^{t_s}| + 2, \ldots, |\Omega_1^{t_s}| + |\Omega_2^{t_s}| + 1 \\ 1, & j = |\Omega_1^{t_s}| + |\Omega_2^{t_s}| + 2 \end{cases} \tag{10}
$$

In our formulation, we have: $\frac{\partial D^{t_{s+1}}(\widetilde{\alpha}^{t_s})}{\partial \widetilde{\alpha}_j^{t_s}} = \frac{1}{n}(\mathbf{c}^{t_j})^T \mathbf{1} - (\widetilde{\mathbf{w}}^{t_s})^T \mathbf{x}_j^{t_{s+1}} = \xi$, if $\widetilde{\alpha}_j^{t_s} \neq 0$, $j = 1, \ldots, |\Omega_1^{t_s}|$, and $\frac{\partial D^{t_{s+1}}(\widetilde{\alpha}^{t_s})}{\partial \widetilde{\alpha}^{t_s}_{|\Omega_1^{t_s}|+1}} = \frac{1}{n}(\mathbf{c}^{t_j})^T \mathbf{1} - (\widetilde{\mathbf{w}}^{t_s})^T \mathbf{x}_j^{t_{s+1}} \ge \xi + \epsilon_1$, $\widetilde{\alpha}_j^{t_s} \neq 0$, and $\frac{\partial D^{t_{s+1}}(\widetilde{\alpha}^{t_s})}{\partial \widetilde{\alpha}_j} = -(\widetilde{\mathbf{w}}^{t_s})^T \mathbf{x}_j^{t_{s+1}} = \zeta$, $j = |\Omega_1^{t_s}| + 2, \ldots, |\Omega_1^{t_s}| + |\Omega_2^{t_s}| + 1$, and

$\frac{\partial D^{t_{s+1}}(\widetilde{\alpha}^{t_s})}{\partial \widetilde{\alpha}^{t_s}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+2}} = -(\widetilde{\mathbf{w}}^{t_s})^T \mathbf{x}^{t_{s+1}}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+2} \geq (\zeta + \epsilon_2)$. Then, we can get:

$$\nabla D^{t_{s+1}}(\widetilde{\alpha}^{t_s})^T \eta \geq \xi + \epsilon_1 + \zeta + \epsilon_2 - \sum_{j=|\Omega_1^{t_s}|+2}^{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+1} \frac{\widetilde{\alpha}_j^{t_s}}{\mu}\zeta - \sum_{j=1}^{|\Omega_1^{t_s}|} \frac{\widetilde{\alpha}_j^{t_s}}{C}\xi \geq \epsilon_1 + \epsilon_2 \geq 0$$

$$\begin{aligned}
\eta^T \mathbf{K} \eta =& (\mathbf{x}^{t_{s+1}}_{|\Omega_1^{t_s}|+1})^2 + (\mathbf{x}^{t_{s+1}}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+2})^2 + \frac{1}{C^2}\sum_{i=1}^{|\Omega_1^{t_s}|}\sum_{j=1}^{|\Omega_1^{t_s}|} \tilde{\alpha}_i^{t_s}\tilde{\alpha}_j^{t_s} K(\mathbf{x}_i^{t_{s+1}}, \mathbf{x}_j^{t_{s+1}}) \\
&+ \frac{1}{\mu^2} \sum_{i=|\Omega_1^{t_s}|+2}^{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+1} \sum_{j=|\Omega_1^{t_s}|+2}^{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+1} \tilde{\alpha}_i^{t_s}\tilde{\alpha}_j^{t_s} K(\mathbf{x}_i^{t_{s+1}}, \mathbf{x}_j^{t_{s+1}}) \\
&+ \frac{2}{C\mu} \sum_{i=1}^{|\Omega_1^{t_s}|} \sum_{j=|\Omega_1^{t_s}|+2}^{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+1} \tilde{\alpha}_i^{t_s}\tilde{\alpha}_j^{t_s} K(\mathbf{x}_i^{t_{s+1}}, \mathbf{x}_j^{t_{s+1}}) + 2K(\mathbf{x}^{t_{s+1}}_{|\Omega_1^{t_s}|+1}, , \mathbf{x}^{t_{s+1}}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+2}) \\
&- \frac{2}{C}\sum_{j=1}^{|\Omega_1^{t_s}|} \tilde{\alpha}_j^{t_s} K(\mathbf{x}_j^{t_{s+1}}, \mathbf{x}^{t_{s+1}}_{|\Omega_1^{t_s}|+1}) - \frac{2}{C}\sum_{j=1}^{|\Omega_1^{t_s}|} \tilde{\alpha}_j^{t_s} K(\mathbf{x}_j^{t_{s+1}}, \mathbf{x}^{t_{s+1}}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+2}) \\
&- \frac{2}{\mu} \sum_{j=|\Omega_1^{t_s}|+1}^{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+2} \tilde{\alpha}_j^{t_s} K(\mathbf{x}_j^{t_{s+1}}, \mathbf{x}^{t_{s+1}}_{|\Omega_1^{t_s}|+1}) - \frac{2}{\mu} \sum_{j=|\Omega_1^{t_s}|+1}^{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+2} \tilde{\alpha}_j^{t_s} K(\mathbf{x}_j^{t_{s+1}}, \mathbf{x}^{t_{s+1}}_{|\Omega_1^{t_s}|+|\Omega_2^{t_s}|+2}) \\
\leq& R^2 + 2R^2 + R^2 + 2R^2 + 2\sqrt{2}R^2 + 2\sqrt{2}R^2 + 2R^2 + 2\sqrt{2}R^2 + 2\sqrt{2}R^2 + 4R^2 \\
=& (12 + 8\sqrt{2})R^2 \qquad (11)
\end{aligned}$$

By utilizing Corollary 13 proved in Tsochantaridis et al. (2006), the minimum increase is:

$$\max_{0 \leq \beta \leq \min\{C,\mu\}} D^{t_{s+1}}(\tilde{\alpha}^{t_s} + \beta\eta) - D^{t_{s+1}}(\tilde{\alpha}^{t_s}) \geq \min\{\frac{\min\{C,\mu\}(\epsilon_1 + \epsilon_2)}{2}, \frac{(\epsilon_1+\epsilon_2)^2}{(24+16\sqrt{2})R^2}\} \qquad (12)$$

Consider all of these three cases together, we will get such a conclusion: For each Cutting Plane iteration, the objective function will increase at least $\min\{\frac{C\epsilon_1}{2}, \frac{\epsilon_1^2}{8R^2}, \frac{\mu\epsilon_2}{2}, \frac{\epsilon_2^2}{16R^2}, \frac{\min\{C,\mu\}(\epsilon_1+\epsilon_2)}{2}, \frac{(\epsilon_1+\epsilon_2)^2}{(24+16\sqrt{2})R^2}\} = \min\{\frac{C\epsilon_1}{2}, \frac{\epsilon_1^2}{8R^2}, \frac{\mu\epsilon_2}{2}, \frac{\epsilon_2^2}{16R^2}, \frac{(\epsilon_1+\epsilon_2)^2}{(24+16\sqrt{2})R^2}\}$, and $R^2 = \max_{i,j} \mathbf{B}_{ij}^2$

References

Joachims, T. Training linear SVMs in linear time. In *KDD*, 2006.

Tsochantaridis, I., Joachims, T., Hofmann, T., and Altun, Y. Large margin methods for structured and interdependent output variables. *JMLR*, 2006.