Supplementary Material for Budgeted Semi-supervised Support Vector Machine

The general update rule is

$$w_{t+1} = w_t - \eta_t g_t - Z_l^t \delta_t D(x_l) - Z_u^t \delta_t D(x_u)$$

$$= \frac{t-1}{t} w_t + C \alpha_t y_t \Phi(x_i) + C^\prime \mu_{uvt} \beta_t (\Phi(x_u) - \Phi(x_v)) - Z_l^t \delta_t D(x_l) - Z_u^t \delta_t D(x_u)$$

We do the convergence analysis for BS3VM with the removal strategy. In this case, the difference vector is $D(x) = \Phi(x)$.

Lemma 1. Let us denote $s_t = \sum_{i \in l_t} |\delta_i| + \sum_{i \in u_t} |\delta_i|$. The following statement holds

$$s_T \leq C + 2C', \forall T$$

Proof. According to the update rule, we have

$$s_{t+1} \leq \frac{t-1}{t} s_t + \frac{C |\alpha_t y_t|}{t} + \frac{2C' \mu_{uvt} |\beta_t|}{t}$$

$$\leq \frac{t-1}{t} s_t + \frac{C + 2C'}{t}$$

where $\alpha_t = -l'_o(w_t; x_i, y_i) = 1_{y_i, w_j \Phi(x_i) \leq 1}$, and $\beta_t = -\text{sign}(w_t^T \Phi_{uv})$.

Here we note that we have used the inequality $\mu_{uvt} = e^{-\frac{\|x_u - x_v\|^2}{2\sigma^2}} < 1$.

It follows that

$$t s_{t+1} \leq (t - 1) s_t + C + 2C'$$

Taking sum when $t = 1, \ldots, T - 1$, we gain

$$(T - 1) s_T \leq (T - 1) \left( C + 2C' \right)$$

$$s_T \leq C + 2C'$$

$\square$

Lemma 2. The following statement holds

$$\|w_t\| \leq C + 2C', \forall t$$

Proof. We have

$$w_t = \sum_{x_i \in l_t} \delta_i \Phi(x_i) + \sum_{x_i \in u_t} \delta_i \Phi(x_i)$$

It follows that

$$\|w_t\| \leq \sum_{x_i \in l_t} |\delta_i| \|\Phi(x_i)\| + \sum_{x_i \in u_t} |\delta_i| \|\Phi(x_i)\| \leq \sum_{x_i \in l_t} |\delta_i| + \sum_{x_i \in u_t} |\delta_i| = s_t \leq C + 2C'$$

$\square$

Lemma 3. The following statement holds

$$\|g_t\| \leq G = 2 \left( C + 2C' \right), \forall t$$
Proof. We have

\[ g_t = J_t'(w_t) = w_t - C\alpha ty_i \Phi (x_i) - C' \mu u_v \beta t (\Phi (x_u) - \Phi (x_v)) \]

where \( \alpha_t = -l'_o(w_t; x_i; y_i) = -1_{y_i}w_t\Phi(x_i) \leq 1 \), and \( \beta_t = -\text{sign}(w_t^T \Phi u_v) \).

It follows that

\[ |g_t| \leq \|w_t\| + C|\alpha_t| \|\Phi (x_i)\| + C'|\mu u_v \beta_t| \|\Phi (x_u) - \Phi (x_v)\| \leq 2\left(C + 2C'\right) \]

Lemma 4. Given two positive integer numbers \( m, n \), assume that before removed the coefficient of \( \Phi (x_i) \) is updated \( m \) times via the vertex sampling and \( n \) times via the edge sampling of the spectral graph. We then have

\[ |\delta_{l_t}| \leq \frac{mC + nC'}{t} \]

Proof. We assume that the coefficient of \( \Phi (x_i) \) is updated via the vertex sampling at iterations \( k_1, ..., k_m \). At the iteration \( k_j \), this coefficient is added by the quantity

\[ -C'_o\left(w_{k_j}; x_{ik_j}, y_{ik_j}\right) \]

At the iteration \( t \), the above quantity becomes

\[ \frac{t - 1}{t} \times \frac{t - 2}{t - 1} \times \cdots \times \frac{k_j}{k_j + 1} \times -C'_o\left(w_{k_j}; x_{ik_j}, y_{ik_j}\right) = -C'_o\left(w_{k_j}; x_{ik_j}, y_{ik_j}\right) \]

We further assume that the coefficient of \( \Phi (x_i) \) is updated via the vertex sampling at iterations \( h_1, ..., h_n \). At the iteration \( h_j \), this coefficient is added by the quantity

\[ -C' \mu u_h \beta_j \text{sign}(w_{h_j}^T \Phi u_h) \]

At the iteration \( t \), the above quantity becomes

\[ \frac{t - 1}{t} \times \frac{t - 2}{t - 1} \times \cdots \times \frac{h_j}{h_j + 1} \times -C' \mu u_h \beta_j \text{sign}(w_{h_j}^T \Phi u_h) = -C' \mu u_h \beta_j \text{sign}(w_{h_j}^T \Phi u_h) \]

Therefore, we have the following representation

\[ \delta_{l_t} = -\sum_{j=1}^{m} C'_o\left(w_{k_j}; x_{ik_j}, y_{ik_j}\right) - \sum_{j=1}^{n} C' \mu u_h \beta_j \text{sign}(w_{h_j}^T \Phi u_h) \]

Hence, we gain the conclusion since

\[ |\delta_{l_t}| \leq \frac{\sum_{j=1}^{m} C'_o\left(w_{k_j}; x_{ik_j}, y_{ik_j}\right) + \sum_{j=1}^{n} C' \mu u_h \beta_j \text{sign}(w_{h_j}^T \Phi u_h)}{t} \leq \frac{mC + nC'}{t} \]

Lemma 5. Given a positive integer number \( p \), assume that before removed the coefficient of \( \Phi (x_i) \) is updated \( p \) times via edge sampling. We then have

\[ |\delta_{u_t}| \leq \frac{pC'}{t} \]

Proof. We skip this proof since it is similar to that of Lemma 4.

Lemma 6. We define \( \rho_i = \frac{\delta_{u_i}}{n} = t\delta_i \) and \( h_t = Z_t^l \rho_i D(x_t) + Z_t^n \rho_n D(x_n) \). Then we have

\[ \|h_t\| \leq H = mC + (n + p)C', \forall t \]
Proof. We have
\[ \|h_t\| \leq |\rho_t| \|D(x_t)\| + |\rho_{ut}| \|D(x_{ut})\| \leq |\rho_t| \|\Phi(x_t)\| + |\rho_{ut}| \|\Phi(x_{ut})\| \leq mC + (n + p)C' = H \]
\[ \square \]

Lemma 7. The following statement holds
\[ \mathbb{E} \left[ \|w_t - w^*\|^2 \right]^{1/2} \leq W, \forall t \]
where \( W = H + \sqrt{H^2 + (G + H)^2} \).

Proof. We have
\[ \|w_{t+1} - w^*\|^2 = \|w_t - \eta_t g_t - Z_t^l \delta_t, D(x_t) - Z_t^u \delta_u, D(x_{ut}) - w^*\|^2 \]
\[ = \|w_t - \eta_t g_t - \eta_t h_t - w^*\|^2 = \|w_t - w^*\|^2 + \eta_t^2 \|g_t + h_t\|^2 - 2\eta_t (w_t - w^*)^T g_t - 2\eta_t (w_t - w^*)^T h_t \]
Taking the conditional expectation w.r.t \( W \), we gain
\[ \mathbb{E} \left[ \|w_{t+1} - w^*\|^2 \right] \leq \mathbb{E} \left[ \|w_t - w^*\|^2 \right] + \eta_t^2 (G + H)^2 - 2\eta_t (w_t - w^*)^T \mathbb{E} [g_t] - 2\eta_t (w_t - w^*)^T \mathbb{E} [h_t] \]
\[ \leq \mathbb{E} \left[ \|w_t - w^*\|^2 \right] + \eta_t^2 (G + H)^2 - 2\eta_t \|w_t - w^*\|^2 \mathbb{E} [h_t] \]
Here we note that we have used the following inequality
\[ (w_t - w^*)^T J' (w_t) \geq J (w_t) - J (w^*) + \frac{1}{2} \|w_t - w^*\|^2 \geq \frac{1}{2} \|w_t - w^*\|^2 \]
Taking the expectation again, we achieve
\[ \mathbb{E} \left[ \|w_{t+1} - w^*\|^2 \right] \leq \frac{t-1}{t} \mathbb{E} \left[ \|w_t - w^*\|^2 \right] + \eta_t^2 (G + H)^2 - 2\eta_t \mathbb{E} [h_t] \]
\[ \leq \frac{t-1}{t} \mathbb{E} \left[ \|w_t - w^*\|^2 \right] + \eta_t^2 (G + H)^2 + 2\eta_t \mathbb{E} \left[ \|w_t - w^*\|^2 \right]^{1/2} \mathbb{E} \left[ \|h_t\|^2 \right]^{1/2} \]
\[ \leq \frac{t-1}{t} \mathbb{E} \left[ \|w_t - w^*\|^2 \right] + \frac{(G + H)^2}{t} + \frac{2H \mathbb{E} \left[ \|w_t - w^*\|^2 \right]^{1/2}}{t} \]
Choosing \( W = H + \sqrt{H^2 + (G + H)^2} \), we have the following: if \( \mathbb{E} \left[ \|w_t - w^*\|^2 \right] \leq W^2 \), \( \mathbb{E} \left[ \|w_{t+1} - w^*\|^2 \right] \leq W^2 \).
\[ \square \]

Theorem 8. Let us consider the running of Algorithm 1. The following statement holds
\[ \mathbb{E} [J(w_t)] - J(w^*) \leq \frac{1}{T} \sum_{t=1}^{T} \mathbb{E} [J(w_t)] - J(w^*) \]
\[ \leq \frac{(G + H)^2 \log T + 1}{2T} + \frac{W}{T} \sum_{t=1}^{T} \mathbb{P}(Z_t^u = 1) \mathbb{E} [\rho_t^2]^{1/2} \]
\[ + \frac{W}{T} \sum_{t=1}^{T} \mathbb{P}(Z_t^u = 1) \mathbb{E} [\rho_{ut}^2]^{1/2} \]
where \( \rho_t = \frac{\delta_t}{m} \) and \( \rho_{ut} = \frac{\delta_{ut}}{m} \).

Proof. We have
\[ \|w_{t+1} - w^*\|^2 = \|w_t - \eta_t g_t - Z_t^l \delta_t, D(x_t) - Z_t^u \delta_u, D(x_{ut}) - w^*\|^2 \]
\[ = \|w_t - \eta_t g_t - \eta_t h_t - w^*\|^2 = \|w_t - w^*\|^2 + \eta_t^2 \|g_t + h_t\|^2 - 2\eta_t (w_t - w^*)^T g_t - 2\eta_t (w_t - w^*)^T h_t \]
\[
(w_t - w^*)^T g_t = \frac{\|w_t - w^*\|^2 - \|w_{t+1} - w^*\|^2}{2\eta_t} + \eta_t \|g_t + h_t\|^2 \frac{2}{\eta_t} - (w_t - w^*)^T h_t
\]

Taking the conditional expectation w.r.t \( t \), we gain

\[
(w_t - w^*)^T \mathbb{E} \left[ J' (w_t) \right] \leq \frac{t - 1}{2} \mathbb{E} \left[ \|w_t - w^*\|^2 \right] - \frac{t - 1}{2} \mathbb{E} \left[ \|w_{t+1} - w^*\|^2 \right] + \frac{(G + H)^2}{2t} - \mathbb{P} (Z_t^1 = 1) \mathbb{E} \left[ (w_t - w^*)^T \rho_t \Phi (x_t) \right] - \mathbb{P} (Z_t^u = 1) \mathbb{E} \left[ (w_t - w^*)^T \rho_u \Phi (x_u) \right]
\]

Taking the expectation of two sides of the above inequality, we obtain

\[
\mathbb{E} \left[ J (w_t) - J (w^*) \right] \leq \frac{t - 1}{2} \mathbb{E} \left[ \|w_t - w^*\|^2 \right] - \frac{t - 1}{2} \mathbb{E} \left[ \|w_{t+1} - w^*\|^2 \right] + \frac{(G + H)^2}{2t} + \mathbb{P} (Z_t^1 = 1) \mathbb{E} \left[ \|w_t - w^*\|^2 \right]^{1/2} \mathbb{E} \left[ \|\rho_t \Phi (x_t)\|^2 \right]^{1/2} + \mathbb{P} (Z_t^u = 1) \mathbb{E} \left[ \|w_t - w^*\|^2 \right]^{1/2} \mathbb{E} \left[ \|\rho_u \Phi (x_u)\|^2 \right]^{1/2}
\]

Taking sum when \( t = 1, ..., T \), we gain

\[
\sum_{t=1}^{T} \mathbb{E} \left[ J (w_t) - J (w^*) \right] \leq \frac{(G + H)^2}{2} \sum_{t=1}^{T} \mathbb{P} (Z_t^1 = 1) \mathbb{E} \left[ \rho_t^2 \right]^{1/2} + \mathbb{W} \sum_{t=1}^{T} \mathbb{P} (Z_t^u = 1) \mathbb{E} \left[ \rho_u^2 \right]^{1/2}
\]

Here we note that \( \rho_t \) and \( \rho_u \) are functionally dependent on \( w_1, ..., w_t, x_1, ..., x_t \) and \( J (w) \) is 1-strongly convex function.