Supplementary Material for Budgeted Semi-supervised Support Vector Machine

The general update rule is

$$\begin{aligned}\mathbf{w}_{t+1} &= \mathbf{w}_t - \eta_t g_t - Z_t^l \delta_{l_t} D\left(x_{l_t}\right) - Z_t^u \delta_{u_t} D\left(x_{u_t}\right) \\ &= \frac{t-1}{t}\mathbf{w}_t + \frac{C\alpha_t y_{i_t}}{t}\Phi\left(x_{i_t}\right) + C'\frac{\mu_{u_t v_t}}{t}\beta_t\left(\Phi\left(x_{u_t}\right) - \Phi\left(x_{v_t}\right)\right) - Z_t^l \delta_{l_t} D\left(x_{l_t}\right) - Z_t^u \delta_{u_t} D\left(x_{u_t}\right)\end{aligned}$$

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

Lemma 1. *Let us denote $s_t = \sum_{i \in I_t^l} |\delta_i| + \sum_{i \in I_t^u} |\delta_i|$. The following statement holds*

$$s_T \leq C + 2C^{'}, \forall T$$

Proof. According to the update rule, we have

$$\begin{aligned}s_{t+1} &\leq \frac{t-1}{t}s_t + \frac{C\left|\alpha_t y_{i_t}\right|}{t} + \frac{2C^{'}\mu_{u_t v_t}\left|\beta_t\right|}{t} \\ &\leq \frac{t-1}{t}s_t + \frac{C+2C^{'}}{t}\end{aligned}$$

where $\alpha_t = -l_o^{'}\left(\mathbf{w}_t; x_{i_t}, y_{i_t}\right) = \mathbb{I}_{y_{i_t}\mathbf{w}_t^{\mathsf{T}}\Phi\left(x_{i_t}\right)\leq 1}$, and $\beta_t = -\text{sign}\left(\mathbf{w}_t^{\mathsf{T}}\mathbf{\Phi}_{u_t v_t}\right)$.

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

It follows that

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

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

$$\begin{aligned}\left(T-1\right)s_T &\leq \left(T-1\right)\left(C+2C^{'}\right) \\ s_T &\leq C+2C^{'}\end{aligned}$$

□

Lemma 2. *The following statement holds*

$$\|\mathbf{w}_t\| \leq C + 2C^{'}, \forall t$$

Proof. We have

$$\mathbf{w}_t = \sum_{i \in I_t^l} \delta_i \Phi\left(x_i\right) + \sum_{i \in I_t^u} \delta_i \Phi\left(x_i\right)$$

It follows that

$$\|\mathbf{w}_t\| \leq \sum_{i \in I_t^l} |\delta_i| \left\|\Phi\left(x_i\right)\right\| + \sum_{i \in I_t^u} |\delta_i| \left\|\Phi\left(x_i\right)\right\| \leq \sum_{i \in I_t^l} |\delta_i| + \sum_{i \in I_t^u} |\delta_i| = s_t \leq C + 2C^{'}$$

□

Lemma 3. *The following statement holds*

$$\|g_t\| \leq G = 2\left(C + 2C^{'}\right), \forall t$$

Proof. We have

$$g_t = \mathcal{J}_t^{'}(\mathbf{w}_t) = \mathbf{w}_t - C\alpha_t y_{i_t}\Phi(x_{i_t}) - C^{'}\mu_{u_t v_t}\beta_t(\Phi(x_{u_t}) - \Phi(x_{v_t}))$$

where $\alpha_t = -l_o^{'}(\mathbf{w}_t; x_{i_t}, y_{i_t}) = -\mathbb{I}_{y_{i_t}\mathbf{w}_t^{\mathsf{T}}\Phi(x_{i_t})\le 1}$, and $\beta_t = -\text{sign}\left(\mathbf{w}_t^{\mathsf{T}}\mathbf{\Phi}_{u_t v_t}\right)$.

It follows that

$$\|g_t\| \le \|\mathbf{w}_t\| + C|\alpha_t|\,\|\Phi(x_{i_t})\| + C^{'}|\mu_{u_t v_t}\beta_t|\,\|\Phi(x_{u_t}) - \Phi(x_{v_t})\| \le 2\left(C + 2C^{'}\right)$$

□

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

$$|\delta_{l_t}| \le \frac{mC + nC^{'}}{t}$$

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

$$\frac{-Cl_o^{'}\left(\mathbf{w}_{k_j}; x_{i_{k_j}}, y_{i_{k_j}}\right)}{k_j}$$

At the iteration t, the above quantity becomes

$$\frac{t-1}{t} \times \frac{t-2}{t-1} \times ... \times \frac{k_j}{k_j+1} \times \frac{-Cl_o^{'}\left(\mathbf{w}_{k_j}; x_{i_{k_j}}, y_{i_{k_j}}\right)}{k_j} = \frac{-Cl_o^{'}\left(\mathbf{w}_{k_j}; x_{i_{k_j}}, y_{i_{k_j}}\right)}{t}$$

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

$$\frac{-C^{'}\mu_{u_{h_j}v_{h_j}}\text{sign}\left(\mathbf{w}_{h_j}^{\mathsf{T}}\Phi_{u_{hj}v_{hj}}\right)}{h_j}$$

At the iteration t, the above quantity becomes

$$\frac{t-1}{t} \times \frac{t-2}{t-1} \times ... \times \frac{h_j}{h_j+1} \times \frac{-C^{'}\mu_{u_{h_j}v_{h_j}}\text{sign}\left(\mathbf{w}_{h_j}^{\mathsf{T}}\Phi_{u_{h_j}v_{h_j}}\right)}{h_j} = \frac{-C^{'}\mu_{u_{h_j}v_{h_j}}\text{sign}\left(\mathbf{w}_{h_j}^{\mathsf{T}}\Phi_{u_{h_j}v_{h_j}}\right)}{t}$$

Therefore, we have the following representation

$$\delta_{l_t} = \frac{-\sum_{j=1}^{m}Cl_o^{'}\left(\mathbf{w}_{k_j}; x_{i_{k_j}}, y_{i_{k_j}}\right) - \sum_{j=1}^{n}C^{'}\mu_{u_{h_j}v_{h_j}}\text{sign}\left(\mathbf{w}_{h_j}^{\mathsf{T}}\Phi_{u_{h_j}v_{h_j}}\right)}{t}$$

Hence, we gain the conclusion since

$$|\delta_{l_t}| \le \frac{\sum_{j=1}^{m}\left|Cl_o^{'}\left(\mathbf{w}_{k_j}; x_{i_{k_j}}, y_{i_{k_j}}\right)\right| + \sum_{j=1}^{n}\left|C^{'}\mu_{u_{h_j}v_{h_j}}\text{sign}\left(\mathbf{w}_{h_j}^{\mathsf{T}}\Phi_{u_{h_j}v_{h_j}}\right)\right|}{t} \le \frac{mC + nC^{'}}{t}$$

□

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

$$|\delta_{u_t}| \le \frac{pC^{'}}{t}$$

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

Lemma 6. *We define $\rho_i = \frac{\delta_i}{\eta_t} = t\delta_i$ and $h_t = Z_t^l\rho_{l_t}D(x_{l_t}) + Z_t^u\rho_{u_t}D(x_{u_t})$. Then we have*

$$\|h_t\| \le H = mC + (n+p)C^{'},\ \forall t$$

Proof. We have

$$\|h_t\| \leq |\rho_{l_t}| \, \|D(x_{l_t})\| + |\rho_{u_t}| \, \|D(x_{u_t})\| \leq |\rho_{l_t}| \, \|\Phi(x_{l_t})\| + |\rho_{u_t}| \, \|\Phi(x_{u_t})\| \leq mC + (n+p) C^{'} = H$$

□

Lemma 7. *The following statement holds*

$$\mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right]^{1/2} \leq W, \forall t$$

where $W = H + \sqrt{H^2 + (G+H)^2}$.

Proof. We have

$$\begin{aligned}\|\mathbf{w}_{t+1} - \mathbf{w}^*\|^2 &= \left\|\mathbf{w}_t - \eta_t g_t - Z_t^l \delta_{l_t} D(x_{l_t}) - Z_t^u \delta_{u_t} D(x_{u_t}) - \mathbf{w}^*\right\|^2 \\ &= \|\mathbf{w}_t - \eta_t g_t - \eta_t h_t - \mathbf{w}^*\|^2 = \|\mathbf{w}_t - \mathbf{w}^*\|^2 + \eta_t^2 \|g_t + h_t\|^2 - 2\eta_t (\mathbf{w}_t - \mathbf{w}^*)^\mathsf{T} g_t - 2\eta_t (\mathbf{w}_t - \mathbf{w}^*)^\mathsf{T} h_t\end{aligned}$$

Taking the conditional expectation w.r.t $\mathbf{w}_t$, we gain

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{w}_{t+1} - \mathbf{w}^*\|^2\right] &\leq \mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right] + \eta_t^2 (G+H)^2 - 2\eta_t (\mathbf{w}_t - \mathbf{w}^*)^\mathsf{T} \mathbb{E}[g_t] - 2\eta_t (\mathbf{w}_t - \mathbf{w}^*)^\mathsf{T} \mathbb{E}[h_t] \\ &\leq \mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right] + \eta_t^2 (G+H)^2 - 2\eta_t (\mathbf{w}_t - \mathbf{w}^*)^\mathsf{T} \mathcal{J}^{'}(\mathbf{w}_t) - 2\eta_t (\mathbf{w}_t - \mathbf{w}^*)^\mathsf{T} \mathbb{E}[h_t] \\ &\leq \mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right] + \eta_t^2 (G+H)^2 - \frac{2\eta_t \|\mathbf{w}_t - \mathbf{w}^*\|^2}{2} - 2\eta_t (\mathbf{w}_t - \mathbf{w}^*)^\mathsf{T} \mathbb{E}[h_t]\end{aligned}$$

Here we note that we have used the following inequality

$$(\mathbf{w}_t - \mathbf{w}^*)^\mathsf{T} \mathcal{J}^{'}(\mathbf{w}_t) \geq \mathcal{J}(\mathbf{w}_t) - \mathcal{J}(\mathbf{w}^*) + \frac{1}{2}\|\mathbf{w}_t - \mathbf{w}^*\|^2 \geq \frac{1}{2}\|\mathbf{w}_t - \mathbf{w}^*\|^2$$

Taking the expectation again, we achieve

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{w}_{t+1} - \mathbf{w}^*\|^2\right] &\leq \frac{t-1}{t}\mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right] + \eta_t^2 (G+H)^2 - 2\eta_t \mathbb{E}\left[(\mathbf{w}_t - \mathbf{w}^*)^\mathsf{T} h_t\right] \\ &\leq \frac{t-1}{t}\mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right] + \eta_t^2 (G+H)^2 + 2\eta_t \mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right]^{1/2} \mathbb{E}\left[\|h_t\|^2\right]^{1/2} \\ &\leq \frac{t-1}{t}\mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right] + \frac{(G+H)^2}{t} + \frac{2H\mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right]^{1/2}}{t}\end{aligned}$$

Choosing $W = H + \sqrt{H^2 + (G+H)^2}$, we have the following: if $\mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right] \leq W^2$, $\mathbb{E}\left[\|\mathbf{w}_{t+1} - \mathbf{w}^*\|^2\right] \leq W^2$. □

Theorem 8. *Let us consider the running of Algorithm 1. The following statement holds*

$$\begin{aligned}\mathbb{E}[\mathcal{J}(\overline{\mathbf{w}}_t)] - \mathcal{J}(\mathbf{w}^*) &\leq \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\mathcal{J}(\mathbf{w}_t)] - \mathcal{J}(\mathbf{w}^*) \\ &\leq \frac{(G+H)^2 (\log T + 1)}{2T} + \frac{W}{T}\sum_{t=1}^{T} \mathbb{P}\left(Z_t^l = 1\right) \mathbb{E}\left[\rho_{l_t}^2\right]^{1/2} \\ &+ \frac{W}{T}\sum_{t=1}^{T} \mathbb{P}\left(Z_t^u = 1\right) \mathbb{E}\left[\rho_{u_t}^2\right]^{1/2}\end{aligned}$$

where $\rho_{l_t} = \frac{\delta_{l_t}}{\eta_t}$ *and* $\rho_{u_t} = \frac{\delta_{u_t}}{\eta_t}$.

Proof. We have

$$\begin{aligned}\|\mathbf{w}_{t+1} - \mathbf{w}^*\|^2 &= \left\|\mathbf{w}_t - \eta_t g_t - Z_t^l \delta_{l_t} D(x_{l_t}) - Z_t^u \delta_{u_t} D(x_{u_t}) - \mathbf{w}^*\right\|^2 \\ &= \|\mathbf{w}_t - \eta_t g_t - \eta_t h_t - \mathbf{w}^*\|^2 = \|\mathbf{w}_t - \mathbf{w}^*\|^2 + \eta_t^2 \|g_t + h_t\|^2 - 2\eta_t (\mathbf{w}_t - \mathbf{w}^*)^\mathsf{T} g_t - 2\eta_t (\mathbf{w}_t - \mathbf{w}^*)^\mathsf{T} h_t\end{aligned}$$

$$(\mathbf{w}_t - \mathbf{w}^*)^\intercal g_t = \frac{\|\mathbf{w}_t - \mathbf{w}^*\|^2 - \|\mathbf{w}_{t+1} - \mathbf{w}^*\|^2}{2\eta_t} + \frac{\eta_t \|g_t + h_t\|^2}{2} - (\mathbf{w}_t - \mathbf{w}^*)^\intercal h_t$$
$$= \frac{\|\mathbf{w}_t - \mathbf{w}^*\|^2 - \|\mathbf{w}_{t+1} - \mathbf{w}^*\|^2}{2\eta_t} + \frac{\eta_t \|g_t + h_t\|^2}{2} - (\mathbf{w}_t - \mathbf{w}^*)^\intercal \left(Z_t^l \rho_{l_t} \Phi(x_{l_t}) + Z_t^u \rho_{u_t} \Phi(x_{u_t})\right)$$

Taking the conditional expectation w.r.t $\mathbf{w}_1, ..., \mathbf{w}_t$, $x_1, \ldots, x_t$ of two sides on the above inequality, we gain

$$(\mathbf{w}_t - \mathbf{w}^*)^\intercal \mathcal{J}'(\mathbf{w}_t) \le \frac{\mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right] - \mathbb{E}\left[\|\mathbf{w}_{t+1} - \mathbf{w}^*\|^2\right]}{2\eta_t} + \frac{\eta_t (G+H)^2}{2}$$
$$- (\mathbf{w}_t - \mathbf{w}^*)^\intercal \left(\mathbb{E}\left[Z_t^l\right] \rho_{l_t} \Phi(x_{l_t}) + \mathbb{E}\left[Z_t^u\right] \rho_{u_t} \Phi(x_{u_t})\right)$$
$$\mathcal{J}(\mathbf{w}_t) - \mathcal{J}(\mathbf{w}^*) + \frac{1}{2}\|\mathbf{w}_t - \mathbf{w}^*\|^2 \le \frac{\mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right] - \mathbb{E}\left[\|\mathbf{w}_{t+1} - \mathbf{w}^*\|^2\right]}{2\eta_t} + \frac{\eta_t (G+H)^2}{2}$$
$$- (\mathbf{w}_t - \mathbf{w}^*)^\intercal \left(\mathbb{P}\left(Z_t^l = 1\right) \rho_{l_t} \Phi(x_{l_t}) + \mathbb{P}\left(Z_t^u = 1\right) \rho_{u_t} \Phi(x_{u_t})\right)$$

Here we note that ρ_{l_t} and ρ_{u_t} are functionally dependent on $\mathbf{w}_1, ..., \mathbf{w}_t$, $x_1, \ldots, x_t$ and $\mathcal{J}(\mathbf{w})$ is 1-strongly convex function. Taking the expectation of two sides of the above inequality, we obtain

$$\mathbb{E}\left[\mathcal{J}(\mathbf{w}_t) - \mathcal{J}(\mathbf{w}^*)\right] \le \frac{t-1}{2}\mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right] - \frac{t}{2}\mathbb{E}\left[\|\mathbf{w}_{t+1} - \mathbf{w}^*\|^2\right] + \frac{(G+H)^2}{2t}$$
$$- \mathbb{P}\left(Z_t^l = 1\right)\mathbb{E}\left[(\mathbf{w}_t - \mathbf{w}^*)^\intercal \rho_{l_t} \Phi(x_{l_t})\right] - \mathbb{P}\left(Z_t^u = 1\right)\mathbb{E}\left[(\mathbf{w}_t - \mathbf{w}^*)^\intercal \rho_{u_t} \Phi(x_{u_t})\right]$$
$$\le \frac{t-1}{2}\mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right] - \frac{t}{2}\mathbb{E}\left[\|\mathbf{w}_{t+1} - \mathbf{w}^*\|^2\right] + \frac{(G+H)^2}{2t}$$
$$+ \mathbb{P}\left(Z_t^l = 1\right)\mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right]^{1/2}\mathbb{E}\left[\|\rho_{l_t}\Phi(x_{l_t})\|^2\right]^{1/2}$$
$$+ \mathbb{P}\left(Z_t^u = 1\right)\mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right]^{1/2}\mathbb{E}\left[\|\rho_{u_t}\Phi(x_{u_t})\|^2\right]^{1/2}$$
$$\le \frac{t-1}{2}\mathbb{E}\left[\|\mathbf{w}_t - \mathbf{w}^*\|^2\right] - \frac{t}{2}\mathbb{E}\left[\|\mathbf{w}_{t+1} - \mathbf{w}^*\|^2\right] + \frac{(G+H)^2}{2t}$$
$$+ \mathbb{P}\left(Z_t^l = 1\right) W \mathbb{E}\left[\rho_{l_t}^2\right]^{1/2} + \mathbb{P}\left(Z_t^u = 1\right) W \mathbb{E}\left[\rho_{u_t}^2\right]^{1/2}$$

Taking sum when $t = 1, ..., T$, we gain

$$\sum_{t=1}^{T}\mathbb{E}\left[\mathcal{J}(\mathbf{w}_t) - \mathcal{J}(\mathbf{w}^*)\right] \le \frac{(G+H)^2}{2}\sum_{t=1}^{T}\frac{1}{t} + W\sum_{t=1}^{T}\mathbb{P}\left(Z_t^l = 1\right)\mathbb{E}\left[\rho_{l_t}^2\right]^{1/2} + W\sum_{t=1}^{T}\mathbb{P}\left(Z_t^u = 1\right)\mathbb{E}\left[\rho_{u_t}^2\right]^{1/2}$$
$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\mathcal{J}(\mathbf{w}_t)\right] - \mathcal{J}(\mathbf{w}^*) \le \frac{(G+H)^2(\log T + 1)}{2T} + \frac{W}{T}\sum_{t=1}^{T}\mathbb{P}\left(Z_t^l = 1\right)\mathbb{E}\left[\rho_{l_t}^2\right]^{1/2} + \frac{W}{T}\sum_{t=1}^{T}\mathbb{P}\left(Z_t^u = 1\right)\mathbb{E}\left[\rho_{u_t}^2\right]^{1/2}$$
$$\mathbb{E}\left[\mathcal{J}(\overline{\mathbf{w}}_t)\right] - \mathcal{J}(\mathbf{w}^*) \le \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\mathcal{J}(\mathbf{w}_t)\right] - \mathcal{J}(\mathbf{w}^*) \le \frac{(G+H)^2(\log T + 1)}{2T} + \frac{W}{T}\sum_{t=1}^{T}\mathbb{P}\left(Z_t^l = 1\right)\mathbb{E}\left[\rho_{l_t}^2\right]^{1/2}$$
$$+ \frac{W}{T}\sum_{t=1}^{T}\mathbb{P}\left(Z_t^u = 1\right)\mathbb{E}\left[\rho_{u_t}^2\right]^{1/2}$$

□