A. Proofs

A.1. Proof of Lemma 1

Proof. Since $P_\lambda(w)$ is λ-strongly convex, $\forall w_1, w_2 \in \mathrm{dom}P_\lambda$,

$$P_\lambda(w_1) \geq P_\lambda(w_2) + g_{P_\lambda}(w_2)^\top(w_1 - w_2) + \frac{\lambda}{2}\|w_1 - w_2\|_2^2,$$

where, $g_{P_\lambda}(w) \in \partial P_\lambda(w)$. On the other hand, $\forall \hat{w} \in \mathrm{dom}P_\lambda$, $g_{P_\lambda}(w^*)^\top(\hat{w} - w^*) \geq 0$ (see Proposition B.24 in (Bertsekas, 1999)). Also, from weak duality, $\forall \hat{\alpha} \in \mathrm{dom}D_\lambda$, $D(\hat{\alpha}) \leq P_\lambda(w^*)$. By substituting $w_1 = \hat{w}, w_2 = w^*$,

$$\frac{\lambda}{2}\|\hat{w} - w^*\|_2^2 \leq P_\lambda(\hat{w}) - D_\lambda(\hat{\alpha}).$$

Therefore, w^* is within a region Θ_{w^*}, where

$$\Theta_{w^*} := \left\{ w \mid \|\hat{w} - w\|_2 \leq \sqrt{2G_\lambda(\hat{w}, \hat{\alpha})/\lambda} \right\}.$$

Since Θ_{w^*} is Sphere, a lower bound of $x_i^\top w^*$ and an upper bound of $x_i^\top w^*$ are given in closed form as follows:

$$LB(x_i^\top w^*) = x_i^\top \hat{w} - \|x_i\|_2\sqrt{2G_\lambda(\hat{w}, \hat{\alpha})/\lambda},$$
$$UB(x_i^\top w^*) = x_i^\top \hat{w} + \|x_i\|_2\sqrt{2G_\lambda(\hat{w}, \hat{\alpha})/\lambda}.$$

■

A.2. Proof of Theorem 2

Proof. Supposing that the result of the previous safe sample screening step assures the optimal values α_i^* for a subset of the samples $i \in \mathcal{S} \subset [n]$, the dual optimal solution region is written as

$$\tilde{\Theta}_{\alpha^*} := \{ \alpha \in \Theta_{\alpha^*} \mid \alpha_i = \alpha_i^* \; \forall i \in \mathcal{S} \}.$$

Then, $X_{:j}^\top \alpha^*$ is bounded from above by the following upper bound:

$$\begin{aligned}
&\tilde{UB}(X_{:j}^\top \alpha^*) \\
&:= \max_{\alpha \in \tilde{\Theta}_{\alpha^*}} X_{:j}^\top \alpha \\
&= \sum_{i \in \mathcal{S}} \alpha_i^* X_{ij} + \max_{\alpha_{\mathcal{U}_s}} X_{\mathcal{U}_s j}^\top \alpha_{\mathcal{U}_s} \\
&\qquad \text{s.t } \|\hat{\alpha}_{\mathcal{U}_s} - \alpha_{\mathcal{U}_s}\|_2^2 \leq r_D^2 - \|\hat{\alpha}_{\mathcal{S}} - \alpha_{\mathcal{S}}^*\|_2^2 \\
&= \sum_{i \in \mathcal{S}} \alpha_i^* X_{ij} + X_{\mathcal{U}_s j}^\top \hat{\alpha}_{\mathcal{U}_s} + \|X_{\mathcal{U}_s j}\|_2\sqrt{r_D^2 - \|\hat{\alpha}_{\mathcal{S}} - \alpha_{\mathcal{S}}^*\|_2^2} \\
&= X_{:j}^\top \tilde{\alpha} + \|X_{\mathcal{U}_s j}\|_2\sqrt{r_D^2 - \|\hat{\alpha}_{\mathcal{S}} - \alpha_{\mathcal{S}}^*\|_2^2}.
\end{aligned}$$

Similarly, $X_{:j}^\top \alpha^*$ is bounded from below by the following lower bound:

$$\tilde{LB}(X_{:j}^\top \alpha) := X_{:j}^\top \tilde{\alpha} - \|X_{\mathcal{U}_s j}\|_2\sqrt{r_D^2 - \|\hat{\alpha}_{\mathcal{S}} - \alpha_{\mathcal{S}}^*\|_2^2}.$$

Therefore,

$$\tilde{UB}(|X_{:j}^\top \alpha|) = |X_{:j}^\top \tilde{\alpha}| + \|X_{\mathcal{U}_s j}\|_2\sqrt{r_D^2 - \|\hat{\alpha}_{\mathcal{S}} - \alpha_{\mathcal{S}}^*\|_2^2}.$$

Since $\tilde{\Theta}_{\alpha^*} \subset \Theta_{\alpha^*}$, the upper bound in (16) is tighter than or equal to that in (9), i.e., $\tilde{UB}(|X_{:j}^\top \alpha^*|) \leq UB(|X_{:j}^\top \alpha^*|)$. ■

A.3. Proof of Theorem 3

Proof. Supposing that the result of the previous safe feature screening step assures that $w_j^* = 0$ for a subset of the features $j \in \mathcal{F} \subset [d]$, the primal optimal solution region is written as

$$\tilde{\Theta}_{w^*} := \{ w \in \Theta_{w^*} \mid w_j = 0 \; \forall j \in \mathcal{F} \}.$$

Then, $x_i^\top w^*$ is bounded from below by the following lower bound:

$$\begin{aligned}
&\tilde{LB}(x_i^\top w) \\
&:= \min_{w \in \tilde{\Theta}_{w^*}} x_i^\top w \\
&= \min_w x_i^\top w \;\text{ s.t. } \|\hat{w} - w\|_2^2 \leq r_P^2, \hat{w}_j = 0 \; \forall j \in \mathcal{F} \\
&= \min_w x_{i\mathcal{U}_f}^\top w_{\mathcal{U}_f} \;\text{ s.t. } \|\hat{w}_{\mathcal{U}_f} - w_{\mathcal{U}_f}\|_2^2 \leq r_P^2 - \|\hat{w}_{\mathcal{F}}\|_2^2 \\
&= x_{i\mathcal{U}_f}^\top \hat{w}_{\mathcal{U}_f} - \|x_{i\mathcal{U}_f}\|_2\sqrt{r_P^2 - \|\hat{w}_{\mathcal{F}}\|_2^2}.
\end{aligned}$$

Similarly, $x_i^\top w^*$ is bounded from above by the following upper bound:

$$UB(x_i^\top w^*) = x_{i\mathcal{U}_f}^\top \hat{w}_{\mathcal{U}_f} + \|x_{i\mathcal{U}_f}\|_2\sqrt{r_P^2 - \|\hat{w}_{\mathcal{F}}\|_2^2}.$$

Since $\tilde{\Theta}_{w^*} \subset \Theta_{w^*}$, these bounds in (17) are tighter than or equal to those in (15), i.e., $\tilde{LB}(x_i^\top w^*) \geq LB(x_i^\top w^*)$ and $\tilde{UB}(x_i^\top w^*) \leq UB(x_i^\top w^*)$. ■

A.4. Proof of Theorem 6

The convex conjugate functions of L_1-penalty and vanilla hinge loss are respectively written as

$$\psi^*(v) := \begin{cases} 0 & (\|v\|_\infty \leq 1), \\ \infty & (\text{otherwise}), \end{cases} \tag{19}$$

$$\ell_i^*(\alpha_i) := \begin{cases} y_i\alpha_i & y_i\alpha_i \in [-1, 0], \\ \infty & (\text{otherwise}), \end{cases} \tag{20}$$

and the dual problem is written as

$$\max_\alpha D_\lambda(\alpha) := \max_\alpha \{y^\top \alpha\}$$
$$\text{s.t. } \left\|\frac{1}{\lambda n}\alpha_i x_i\right\|_\infty \leq 1, \; y_i\alpha_i \in [0, 1] \;\; \forall i \in [n].$$

We first construct the the dual optimal solution region $\tilde{\Theta}_{\alpha^*}$.

Lemma 8. *For an arbitrary pair of primal feasible solution $\hat{w} \in \mathrm{dom}P_\lambda$ and dual feasible solution $\hat{\alpha} \in \mathrm{dom}D_\lambda$, the dual optimal solution region is written as*

$$\Theta_{\alpha^*} := \left\{ \forall i\ y_i\alpha_i \in [0,1] \;\middle|\; y^\top\hat{\alpha} \le y^\top\alpha \le P_\lambda(\hat{w}) \right\}.$$

Proof of Lemma 8 . From the optimality and weak duality $y^\top\hat{\alpha} \le y^\top\alpha^*$ and $y^\top\alpha^* \le P_\lambda(\hat{w})$, respectively. Therefore,

$$\alpha^* \in \hat{\Theta}_{\alpha^*} := \left\{ \alpha \in \mathrm{dom}D_\lambda \;\middle|\; y^\top\hat{\alpha} \le y^\top\alpha \le P_\lambda(\hat{w}) \right\}.$$

Noting that $\hat{\Theta}_{\alpha^*} \subseteq \Theta_{\alpha^*}$, $\alpha^* \in \Theta_{\alpha^*}$. ■

Proof of Theorem 6. From Lemma 8,

$$X_{:j}^\top\alpha^* \ge LB(X_{:j}^\top\alpha^*) := \min_{\alpha\in\Theta_{\alpha^*}} X_{:j}^\top\alpha$$

Moreover,

$$LB(X_{:j}^\top\alpha^*) = \min_{\alpha\in\Theta_{\alpha^*}} Z_{:j}^\top\alpha_y,$$

where $\alpha_y := [y_1\alpha_1, \dots, y_n\alpha_n]^\top$. Let us define three n-dimensional vectors $\bar{\alpha}^{(1)}$, $\bar{\alpha}^{(2)}$ and $\bar{\alpha}^{(3)}$ as follows:

$$\bar{\alpha}_i^{(1)} := \begin{cases} y_i & (Z'_{ij} < 0) \\ 0 & (\text{otherwise}), \end{cases}$$
$$\bar{\alpha}_i^{(2)} := \begin{cases} y_i & (Z'_{ij} \le Z'_{l_q j}) \\ y_i(y^\top\hat{\alpha} - l_q) & (Z'_{ij} = Z'_{(l_q+1)j}) \\ 0 & (otherwise), \end{cases}$$
$$\bar{\alpha}_i^{(3)} := \begin{cases} y_i & (Z'_{ij} \le Z'_{u_q j}) \\ y_i(P_\lambda(\hat{w}) - u_q) & (Z'_{ij} = Z'_{(u_q+1)j}) \\ 0 & (otherwise). \end{cases}$$

If $l_q + 1 \le n_{Z'_{:j}} \le u_q$, then $\bar{\alpha}^{(1)}$ is an element of Θ_{α^*} and minimizes $X_{:j}^\top\alpha$. If $n_{Z'_{:j}} < l_q + 1$ then $\bar{\alpha}^{(1)} \notin \Theta_{\alpha^*}$, $\bar{\alpha}^{(2)}$ is an element of Θ_{α^*} and minimizes $X_{:j}^\top\alpha$ because $y^\top\bar{\alpha}^{(2)} = y^\top\hat{\alpha}$. If $m_{Z'_{:j}} > u_q$ then $\bar{\alpha}^{(1)} \notin \Theta_{\alpha^*}$, meaning that $\bar{\alpha}^{(3)}$ is an element of Θ_{α^*} and minimizes $X_{:j}^\top\alpha$ because $y^\top\bar{\alpha}^{(3)} = P_\lambda$. Therefore,

$$LB(X_{:j}^\top\alpha^*) := \begin{cases} \sum_{i=1}^{l_q} Z'_{ij} + (y^\top\hat{\alpha} - l_q)Z'_{(l_q+1)j} & (n_{Z'_{:j}} < l_q + 1), \\ \sum_{i=1}^{u_q} Z'_{ij} + (P_\lambda(\hat{w}) - u_q), Z'_{u_q j} & (n_{Z'_{:j}} > u_q) \\ \sum_{i=1}^{n} \min\{0, Z'_{ij}\} & (\text{otherwise}), \end{cases}$$

Similarly, from Lemma 8,

$$X_{:j}^\top\alpha^* \le UB(X_{:j}^\top\alpha^*) := \max_{\alpha\in\Theta_{\alpha^*}} X_{:j}^\top\alpha$$

Moreover,

$$UB(X_{:j}^\top\alpha^*) = \max_{\alpha\in\Theta_{\alpha^*}} Z_{:j}^\top\alpha_y.$$

Let us define three n-dimensional vectors $\bar{\alpha}^{(4)}$, $\bar{\alpha}^{(5)}$ and $\bar{\alpha}^{(6)}$ as follows:

$$\bar{\alpha}_i^{(4)} := \begin{cases} y_i & (Z'_{ij} > 0) \\ 0 & (\text{otherwise}), \end{cases}$$
$$\bar{\alpha}_i^{(5)} := \begin{cases} y_i & (Z'_{ij} \ge Z'_{(n-l_q)j}) \\ y_i(y^\top\hat{\alpha} - l_q) & (Z'_{ij} = Z'_{(n-l_q-1)j}) \\ 0 & (otherwise), \end{cases}$$
$$\bar{\alpha}_i^{(6)} := \begin{cases} y_i & (Z'_{ij} \ge Z'_{(n-u_q)j}) \\ y_i(P_\lambda(\hat{w}) - u_q) & (Z'_{ij} = Z'_{(n-u_q-1)j}) \\ 0 & (otherwise). \end{cases}$$

If $l_q + 1 \le p_{Z'_{:j}} \le u_q$, then $\bar{\alpha}^{(4)}$ is an element of Θ_{α^*} and maximizes $X_{:j}^\top\alpha$. If $p_{Z'_{:j}} < l_q + 1$ then $\bar{\alpha}^{(4)} \notin \Theta_{\alpha^*}$, $\bar{\alpha}^{(5)}$ is an element of Θ_{α^*} and maximizes $X_{:j}^\top\alpha$ because $y^\top\bar{\alpha}^{(5)} = y^\top\hat{\alpha}$. If $p_{Z'_{:j}} > u_q$ then $\bar{\alpha}^{(4)} \notin \Theta_{\alpha^*}$, meaning that $\bar{\alpha}^{(6)}$ is an element of Θ_{α^*} and maximizes $X_{:j}^\top\alpha$ because $y^\top\bar{\alpha}^{(6)} = P_\lambda$. Therefore,

$$UB(X_{:j}^\top\alpha^*) := \begin{cases} \sum_{i=n-l_q}^{n} Z'_{ij} + (y^\top\hat{\alpha} - l_q)Z'_{(n-l_q-1)j} & (p_{Z'_{:j}} < l_q + 1), \\ \sum_{i=n-u_q}^{n} Z'_{ij} + (P_\lambda(\hat{w}) - u_q)Z'_{(n-u_q-1)j} & (p_{Z'_{:j}} > u_q), \\ \sum_{i=1}^{n} \max\{0, Z'_{ij}\} & (\text{otherwise}). \end{cases}$$

On the other hand, from KKT condition(6),

$$\frac{1}{\lambda n} X_{:j}^\top\alpha^* \in \begin{cases} \frac{w_j^*}{|w_j^*|} & (w_j^* \neq 0) \\ [-1, 1] & (\text{otherwise}). \end{cases} \tag{21}$$

Therefore, if $LB(X_{:j}^\top\alpha^*) < -\lambda n$ and $UB(X_{:j}^\top\alpha^*) > \lambda n$ then $w_j^* = 0$. ■

A.5. Proof of Theorem 7

First, we construct the primal optimal solution region Θ_{w^*}.

Lemma 9. *The primal optimal solution region Θ_{w^*} is given $\forall\hat{w} \in \mathrm{dom}P_\lambda$ as*

$$\Theta_{w^*} = \left\{ w \in \mathrm{dom}P_\lambda \;\middle|\; \lambda\|w\|_1 + g_\ell(\hat{w})^\top w \le k \right\}, \tag{22}$$

where $g_\ell(w) := \frac{1}{n}\sum_{i\in[n]} g_{\ell_i}(w)$.

Proof. From Proposition B.24 in (Bertsekas, 1999),

$$(\lambda g_\psi(w^*) + g_\ell(w^*))^\top(w^* - \hat{w}) \le 0, \forall\hat{w} \in \mathrm{dom}P_\lambda,$$

where $g_\psi(w) \in \partial\psi(w)$. Form the convexity of ℓ_i for $i \in [n]$ and the definition of subgradient

$$\ell_i(w^*) \ge \ell_i(\hat{w}) + g_{\ell_i}(\hat{w})(w^* - \hat{w}), \forall\hat{w} \in \mathrm{dom}P_\lambda$$
$$\ell_i(\hat{w}) \ge \ell_i(w^*) + g_{\ell_i}(w^*)(\hat{w} - w^*), \forall\hat{w} \in \mathrm{dom}P_\lambda,$$

and thus, $g_{\ell_i}(w^*)^\top(w^* - \hat{w}) \geq g_{\ell_i}(\hat{w})^\top(w^* - \hat{w}), \forall \hat{w} \in \mathrm{dom}P_\lambda$. Therefore, $\forall \hat{w} \in \mathrm{dom}P_\lambda$,

$$\lambda g_\psi(w^*)^\top w^* + g_\ell(\hat{w})^\top w^* \leq \lambda g_\psi(w^*)^\top \hat{w} + g_\ell(\hat{w})^\top \hat{w}.$$

Since $g_\psi(\hat{w})^\top \hat{w} = \|\hat{w}\|_1 = \max_{s \in [-1,1]^d} s^\top \hat{w}$ and $g_\psi(w^*) \in [-1,1]^d$, we have

$$\lambda g_\psi(w^*)^\top \hat{w} \leq \lambda g_\psi(\hat{w})^\top \hat{w}.$$

By combining these results,

$$\lambda \|w^*\|_1 + g_\ell(\hat{w})^\top w^* \leq k, \ \ \forall \hat{w} \in \mathrm{dom}P_\lambda.$$

■

Proof of Theorem 7. From Lemma 9,

$$x_i^\top w^* \geq LB(x_i^\top w^*) := \min_{w \in \Theta_{w^*}} x_i^\top w.$$

Using a Lagrange multiplier $\mu > 0$,

$$\begin{aligned} LB(x_i^\top w^*) &= \min x_i^\top w \ \text{ s.t } \ w \in \Theta_{w^*} \qquad (23) \\ &= \min_w \max_{\mu>0} \{x_i^\top w + \mu(\lambda\|w\|_1 + g_\ell(\hat{w})^\top w - k)\} \\ &= \max_{\mu>0} \{\mu k + \min(\underbrace{x_i^\top w + \mu\lambda\|w\|_1 + \mu g_\ell(\hat{w})^\top w}_{L(w)})\} \end{aligned}$$

Since $0 \in \partial L$, which is written as $\partial L = x_i + \mu\lambda\partial\psi(w) + \mu g_\ell(\hat{w})$, we have

$$\mu\lambda g_\psi(w) = -x_i - \mu g_\ell(\hat{w}) \qquad (24)$$

Substituting $\mu\lambda\|w\| = -x_i^\top w - \mu g_\ell(\hat{w})^\top w$ into (23),

$$\begin{aligned} LB(x_i^\top w^*) &= \max_{\mu>0}\{\mu k\} \\ \text{s.t. } & \| -\frac{1}{\lambda} x_i^\top w - \frac{\mu}{\lambda} g_\ell(\hat{w})^\top w\|_\infty \leq \mu, \end{aligned}$$

where the constraint comes from (24). Similarly, since

$$x_i^\top w^* \leq UB(x_i^\top w^*) := \max_{w \in \Theta_{w^*}} x_i^\top w = -\min_{w \in \Theta_{w^*}} x_i^\top w,$$

$$\begin{aligned} UB(x_i^\top w^*) &= \max_{\mu>0}\{\mu k\} \\ \text{s.t. } & \|\frac{1}{\lambda} x_i^\top w - \frac{\mu}{\lambda} g_\ell(\hat{w})^\top w\|_\infty \leq \mu. \end{aligned}$$

■

B. Safe keeping by using KKT optimality conditions

In this appendix, we describe another type of safe keeping approaches based on KKT optimality conditions.

Theorem 10. *For an arbitrary pair of primal feasible solution $\hat{w} \in \mathrm{dom}P_\lambda$ and dual feasible solution $\hat{\alpha} \in \mathrm{dom}D_\lambda$,*

$$LB(X_{:j}^\top \alpha^*) < -\lambda n \text{ and } \lambda n < UB(X_{:j}^\top \alpha^*) \ \Rightarrow \ w_j^* \neq 0$$

for $j \in [d]$, where

$$\begin{aligned} LB(X_{:j}^\top \alpha^*) &:= X_{:j}^\top \hat{\alpha} - \|X_{:j}\|_2 \sqrt{2nG_\lambda(\hat{w}, \hat{\alpha})/\gamma}, \\ UB(X_{:j}^\top \alpha^*) &:= X_{:j}^\top \hat{\alpha} + \|X_{:j}\|_2 \sqrt{2nG_\lambda(\hat{w}, \hat{\alpha})/\gamma}. \end{aligned}$$

Proof. In the case that D_λ is γ/n-strongly concave, $X_{:j}^\top \alpha^*$ is bounded from below and above respectively by the following lower and upper bounds:

$$\begin{aligned} LB(X_{:j}^\top \alpha^*) &:= X_{:j}^\top \hat{\alpha} - \|X_{:j}\|_2 \sqrt{2nG_\lambda(\hat{w}, \hat{\alpha})/\gamma}, \\ UB(X_{:j}^\top \alpha^*) &:= X_{:j}^\top \hat{\alpha} + \|X_{:j}\|_2 \sqrt{2nG_\lambda(\hat{w}, \hat{\alpha})/\gamma}. \end{aligned}$$

On the other hand, in the case of our specific regularization term (2), from KKT optimality condition (6), if $-\lambda n < X_{:j}^\top \alpha^* < \lambda n$ then $w_j^* \neq 0$.

Therefore,

$$LB(X_{:j}^\top \alpha^*) < -\lambda n \text{ and } \lambda n < UB(X_{:j}^\top \alpha^*) \ \Rightarrow \ w_j^* \neq 0$$

■

Similarly, we can develop safe sample keeping based on KKT optimality condition.

Theorem 11. *For an arbitrary pair of primal feasible solution $\hat{w} \in \mathrm{dom}P_\lambda$ and dual feasible solution $\hat{\alpha} \in \mathrm{dom}D_\lambda$, if ℓ_i is smoothed hinge loss then, for $y_i = +1$,*

$$1 - \gamma < LB(x_i^\top w^*) \text{ and } UB(x_i^\top w^*) < 1 \ \Rightarrow \ \alpha_i^* \notin \{0, +1\},$$

and, for $y_i = -1$,

$$-1 < LB(x_i^\top w^*) \text{ and } UB(x_i^\top w^*) < \gamma - 1 \ \Rightarrow \ \alpha_i^* \notin \{-1, 0\}.$$

If ℓ_i is smoothed ε-insensitive loss then

$$\begin{gathered} -\gamma + y_i - \varepsilon < LB(x_i^\top w^*) \text{ and } UB(x_i^\top w^*) < y_i - \varepsilon \\ \text{or} \\ y_i + \varepsilon < LB(x_i^\top w^*) \text{ and } UB(x_i^\top w^*) < \gamma + y_i + \varepsilon \\ \Rightarrow \ \alpha_i^* \notin \{-1, 0, +1\}, \end{gathered}$$

for $j \in [d]$, where

$$\begin{aligned} LB(x_i^\top w^*) &= x_i^\top \hat{w} - \|x_i\|_2 \sqrt{2G_\lambda(\hat{w}, \hat{\alpha})/\lambda}, \\ UB(x_i^\top w^*) &= x_i^\top \hat{w} + \|x_i\|_2 \sqrt{2G_\lambda(\hat{w}, \hat{\alpha})/\lambda}. \end{aligned}$$

Proof. In the case that P_λ is λ-strongly convex, $x_i^\top w^*$ is bounded from below and above respectively by the following lower and upper bounds:

$$LB(x_i^\top w^*) = x_i^\top \hat{w} - \|x_i\|_2 \sqrt{2G_\lambda(\hat{w}, \hat{\alpha})/\lambda},$$
$$UB(x_i^\top w^*) = x_i^\top \hat{w} + \|x_i\|_2 \sqrt{2G_\lambda(\hat{w}, \hat{\alpha})/\lambda}.$$

On the other hand, from KKT optimality condition (6), in the case of smoothed hinge loss (3), if $y_i = +1$ and $1 - \gamma < x_i^\top w^* < 1$ then $\alpha_i^* \in \{0, +1\}$, if $y_i = 1$ and $-1 < x_i^\top w^* < \gamma - 1$ then $\alpha_i^* \in \{-1, 0\}$. Therefore,

$$y_i = +1 \text{ and } 1 - \gamma < LB(x_i^\top w^*) \text{ and } UB(x_i^\top w^*) < 1$$
$$\Rightarrow\ \alpha_i^* \notin \{0, +1\},$$
$$y_i = -1 \text{ and } -1 < LB(x_i^\top w^*) \text{ and } UB(x_i^\top w^*) < \gamma - 1$$
$$\Rightarrow\ \alpha_i^* \notin \{-1, 0\}.$$

Also, in the case of smoothed ε-insensitive (4), if $-\gamma + y_i - \varepsilon < x_i^\top w^* < y_i - \varepsilon$ or $y_i + \varepsilon < x_i^\top w^* < \gamma + y_i + \varepsilon$ then, $\alpha_i^* \notin \{-1, 0, +1\}$. Therefore,

$$-\gamma + y_i - \varepsilon < LB(x_i^\top w^*) \text{ and } UB(x_i^\top w^*) < y_i - \varepsilon$$
$$\text{or}$$
$$y_i + \varepsilon < LB(x_i^\top w^*) \text{ and } UB(x_i^\top w^*) < \gamma + y_i + \varepsilon$$
$$\Rightarrow\ \alpha_i^* \notin \{-1, 0, +1\},$$

■

C. Other experiments

In this appendix, we show the rest of the experimental results.

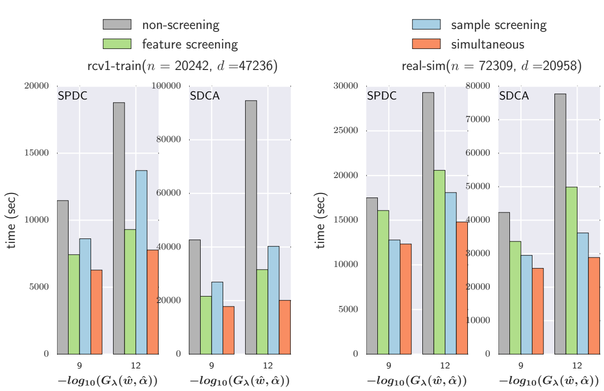

Figure 5. Total computation time for training 100 solutions for various values of λ in regression problems.

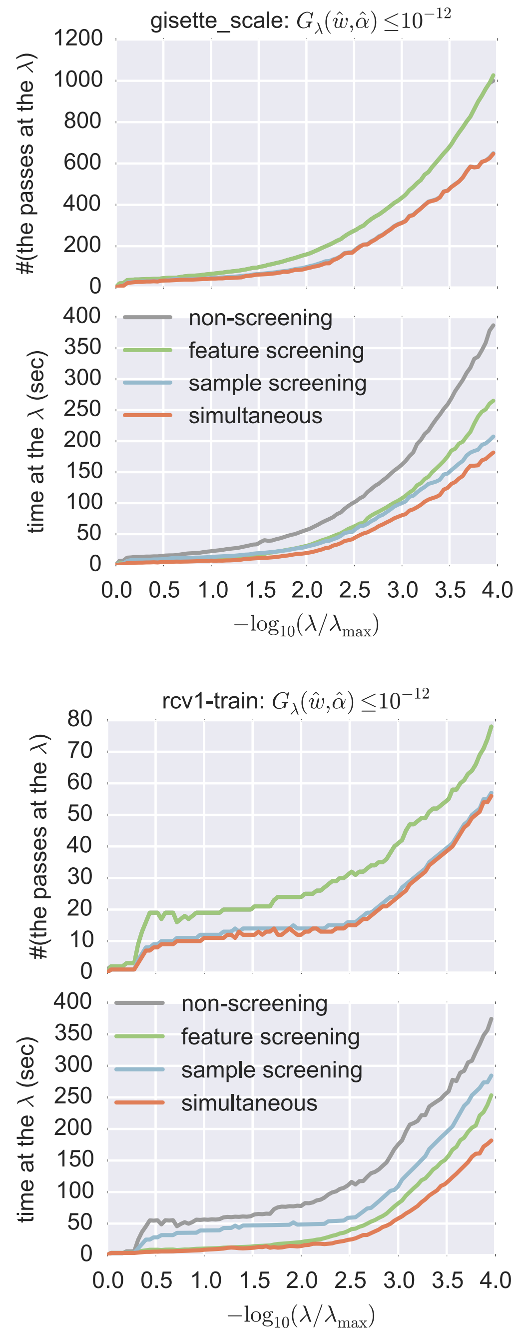

Figure 6. Number of optimization steps and computation time in classification problems (`rcv1-train` and `rcv1-train` datasets). See the caption in Figure 4.

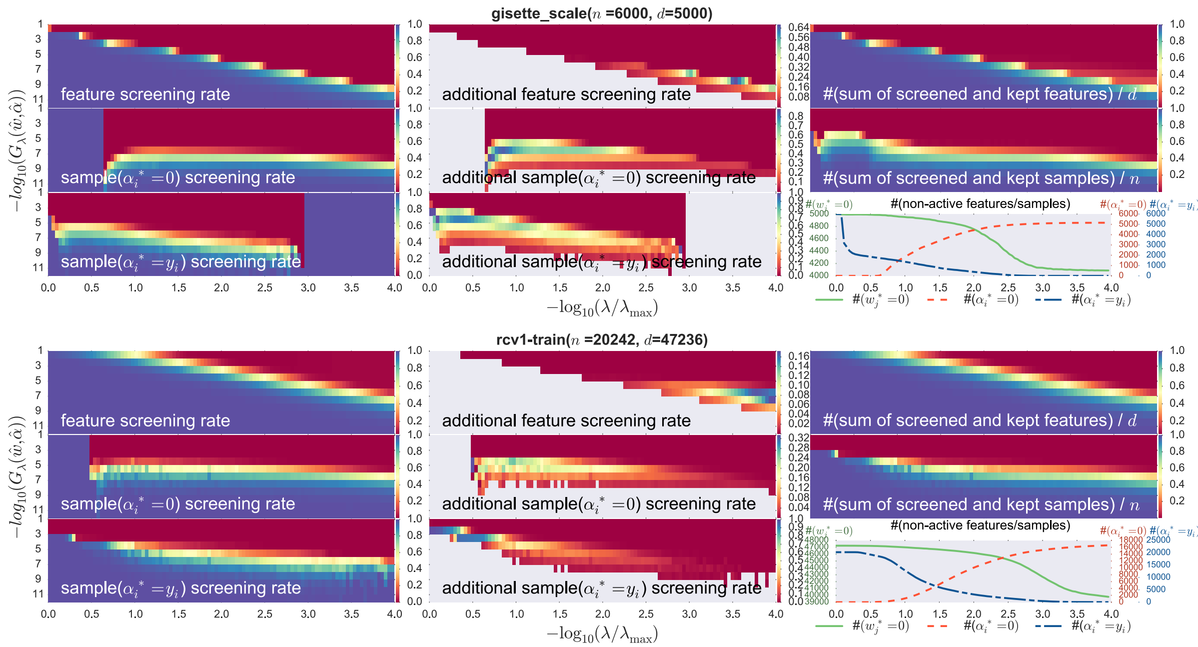

Figure 7. Safe screening and keeping rates in classification problems (`rcv1-train` and `rcv1-train` datasets). See the caption in Figure 2.

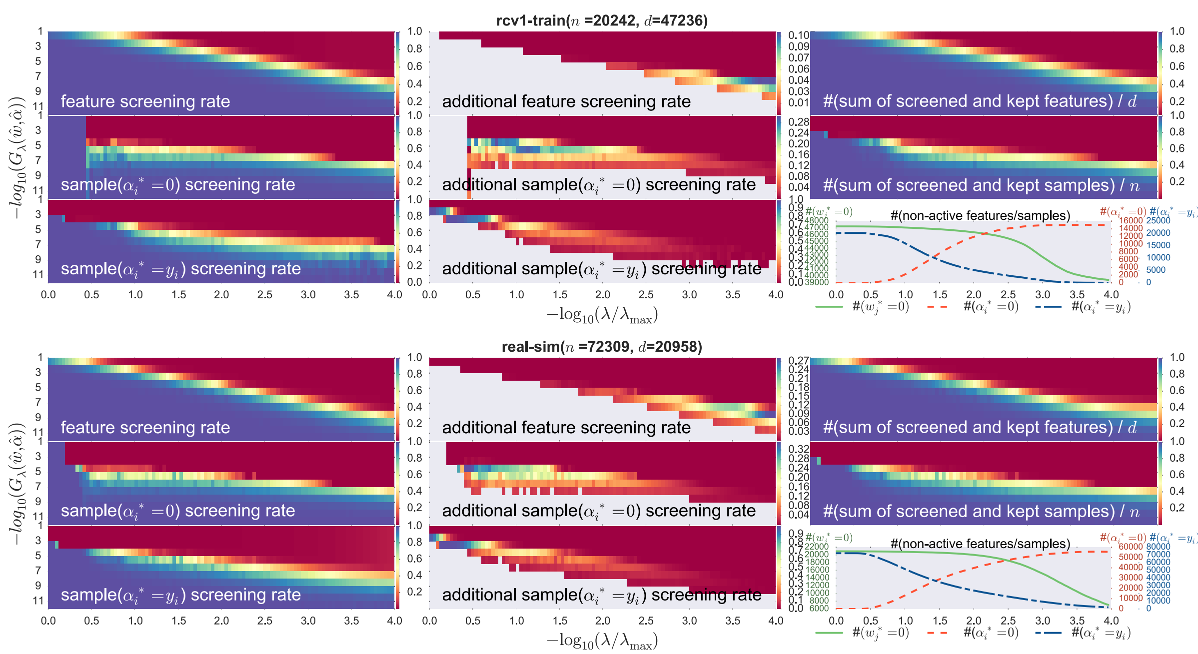

Figure 8. Safe screening and keeping rates in regression problems. See the caption in Figure 2.