Research Write Up

We assume $\exists\ \theta^*$ st $y_i = \text{sign}(\theta^{*T} x_i)$ or $y_i^{(corr)} = \text{sign}(\theta^{*T} x_i + \varepsilon')$
$\varepsilon' \in [-\varepsilon, \varepsilon]$

We have data: $\{(x_1, y_1) \cdots (x_n, y_n)\}$

We have test points: $\{z_1, \ldots, z_m\}$ where $z_i \sim P_t$ is drawn from some test distribution such that $P(\text{sign}(\theta^{*T} z) = 1) = p$ where p is known to us.

We want to minimize:

$$\min_\theta \ \left(|\{z_i \mid \theta^T z_i \geq 0\}| - mp \right)^2$$
$$\text{s.t.} \quad D\theta \leq 0$$

could replace with ε.

where

$$D = -\begin{bmatrix} - y_1 x_1 - \\ \vdots \\ - y_n x_1 - \end{bmatrix}$$

Instead we minimize:

$$\min_\theta \ \left(\sum_{i=1}^m \sigma_t(\theta^T z_i) - mp \right)^2$$
$$\text{s.t.} \quad D\theta \leq 0$$

where $\sigma_t(x) = (1 + \exp(-tx))^{-1}$, and t is a fixed constant > 0.

This is a convex-composite formulation. Let

$f(\theta) = \left(\sum_{i=1}^m \sigma_t(\theta^T z_i) - mp \right)^2$ then $f(\theta) = h(c(\theta))$ where

$c(\theta) = \sum_i \sigma_t(\theta^T z_i) - mp$ and $h(x) = x^2$. Note h is convex and c is smooth.

We can use the convex composite prox-linear algorithm.

Let $\hat{f}_\beta(\theta) := h(c(\beta) + \nabla c(\beta)^T (\theta - \beta))$

$$= \left(\sum_{i=1}^m \sigma_t(\beta^T z_i) - mp + \left[\sum_{i=1}^m \frac{e^{-\beta^T z_i}}{(1 + e^{-\beta^T z_i})^2} z_i \right]^T (\theta - \beta) \right)^2$$

Let $\beta = \theta^{(k)}$, to compute our next iterate we want to solve:

$$\min.\ \hat{f}_\beta(\theta)$$
$$\text{s.t.} \quad D\theta \preceq 0$$

Let $V_\beta := \sum_{i=1}^{m} \frac{e^{-\beta^T z_i}}{(1+e^{-\beta^T z_i})^2} z_i$ and $C_\beta = \sum_{i=1}^{m} \sigma_t(\beta^T z_i) - mp - V_\beta^T \beta$

Then $\hat{f}_\beta(\theta) = (C_\beta + V_\beta^T \theta)^2$

Our problem is convex with an affine inequality constraint so Slater's holds and we consider the KKT conditions:

$$L(\theta, \lambda) = (C_\beta + V_\beta^T \theta)^2 + \lambda^T D\theta$$

KKT conditions:

1. $\frac{\partial L}{\partial \theta}\Big|_{\theta^*} = 0 \Rightarrow 2(C_\beta + V_\beta^T \theta^*) V_\beta + D^T \lambda^* = 0$

2. $\lambda^{*T} D\theta^* = 0$

3. $\lambda^* \geq 0$

4. $D\theta^* \preceq 0$

By 1 we have that $D^T\lambda^* = -2(C_\beta + V_\beta^T \theta^*) V_\beta = \alpha V_\beta$

Plugging into 2 we get $\alpha V_\beta^T \theta^* = 0$ thus we have that either a $\alpha = 0$ or b $V_\beta^T \theta^* = 0$

We can show (refine this) that only a is possible, thus

$$D^T \lambda^* = 0 = -2(C_\beta + V_\beta^T \theta^*) V_\beta$$
$$\Rightarrow -C_\beta = V_\beta^T \theta^*$$

So we want to find θ^* such that

$$V_\beta^T \theta^* = -C_\beta \quad \text{and} \quad D\theta^* \preceq 0$$

Find θ' s.t. $D\theta' \preceq 0$ and we want to find some $s > 0$ such that

$$v_\beta^T(s\,\theta') = -c_\beta$$

$$s = -\frac{c_\beta}{v_\beta^T \theta'} \Rightarrow \text{we need} \quad \begin{matrix} v_\beta^T\theta' > 0 & \text{if } c_\beta < 0 \\ v_\beta^T\theta' < 0 & \text{if } c_\beta > 0 \end{matrix}$$

Note that if there exists a feasible θ^* (i.e. $D\theta^* \preceq 0$) then

$\underset{\theta}{\mathrm{argmin}}\{1^T D\theta\}$ will be feasible. We can restrict the norm of θ since if $\exists$ feasible θ^* then $\frac{\theta^*}{\|\theta^*\|_2^2}$ is also feasible and so a minimizer $\underset{\|\theta\|_2^2 \le 1}{\mathrm{argmin}}\{1^T D\theta\}$ will be feasible. Let

$$\tilde{D} = \begin{bmatrix} D \\ v_\beta^T \end{bmatrix} \quad \text{if } c_\beta > 0$$

or

$$\tilde{D} = \begin{bmatrix} D \\ -v_\beta^T \end{bmatrix} \quad \text{if } c_\beta < 0$$

$$\Bigg\} \quad \tilde{D} = \begin{bmatrix} D \\ \mathrm{sign}(c_\beta)\, v_\beta^T \end{bmatrix}$$

We thus can solve:

$$\min_\theta \; 1^T \tilde{D}\theta$$
$$\text{s.t. } \|\theta\|_2^2 \le 1$$

We have:

$$L(\theta, \lambda) = 1^T\tilde{D}\theta + \lambda(\|\theta\|_2^2 - 1)$$

Again using KKT conditions we get

① $\left.\frac{\partial L}{\partial \theta}\right|_{\theta=\theta'} = \tilde{D}^T 1 + 2\lambda\theta' = 0$

$$\Rightarrow \theta' = \left(-\frac{1}{2\lambda}\right)\tilde{D}^T 1$$

Since the scaling of θ' doesn't matter set $\theta' = -\tilde{D}^T 1$

$$-\tilde{D}^T 1 = \begin{bmatrix} | & & | & | \\ x_1 y_1 & \cdots & x_n y_n & -\mathrm{sign}(c_\beta) v_\beta \\ | & & | & | \end{bmatrix} 1$$

$$\Rightarrow \quad \theta' = \sum_{i=1}^{n} x_i y_i \; - \; \mathrm{sign}(C_\beta) V_\beta$$

$$\theta^* = s\theta' = -\frac{C_\beta}{V_\beta^T \theta'} \theta' = -\frac{C_\beta}{V_\beta^T \left(\sum_i x_i y_i - \mathrm{sign}(C_\beta) V_\beta\right)} \left(\sum_i x_i y_i - \mathrm{sign}(C_\beta) V_\beta\right)$$

where again $\quad C_\beta = \sum_{i=1}^{m} \sigma_t(\beta^T z_i) - m p - V_\beta^T \beta$

and $\quad V_\beta = \sum_{i=1}^{m} \frac{e^{-\beta^T z_i}}{(1+e^{-\beta^T z_i \cdot t})^2} z_i \cdot t$.