- SVM formulations

- $$\inf_{\substack{W\lambda \\ s_i \\ p_i^+, p_i^-}} \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i$$

$$\text{s.t} \quad 1 - \hat{y}_i(\langle w, \hat{x}_i\rangle) + \langle p_i^+, d - C\hat{x}_i\rangle \leq s_i \quad i \in [N]$$

$$1 + \hat{y}_i(\langle w, \hat{x}_i\rangle) + \langle p_i^-, d - C\hat{x}_i\rangle - k\lambda \leq s_i \quad i \in [N]$$

$$\| C^T p_i^+ + \hat{y}_i w \|_* \leq \lambda$$

$$\| C^T p_i^- - \hat{y}_i w \|_* \leq \lambda$$

$$s_i \geq 0$$

$$p_i^+, p_i^- \in \mathcal{C}^*$$

Proof

→ The general formulation for a piecewise affine loss function is given below

$$L(z) = \max_{j \in J} \{ a_j z + b_j \}.$$

$$\inf_{\substack{w, \lambda, s_i \\ p_i^+, p_i^-}} \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i$$

$$\text{s.t} \quad S_X(a_j \hat{y}_i w - p_{ij}^+) + b_j + \langle p_{ij}^+, \hat{x}_i\rangle \leq s_i$$

$$S_X(-a_j \hat{y}_i w - p_{ij}^-) + b_j + \langle p_{ij}^-, \hat{x}_i\rangle - k\lambda \leq s_i \quad i \in [N] \quad j \in [J]$$

$$\| p_{ij}^+ \|_* \leq \lambda$$

$$\| p_{ij}^- \|_* \leq \lambda$$

where S_X denotes the support function of X

For the SVM setting we consider that the input space $\mathbb{X}$ admits a conic representation.

$$\mathbb{X} = \{x \in \mathbb{R}^n, \; Cx \preceq_{\mathcal{C}} d\}$$

for some matrix C, vector d and proper convex cone $\mathcal{C}$ of appropriate dimensions. We also assume that $\mathbb{X}$ admits a slater point $x_s \in \mathbb{R}^n$ with $Cx_s \prec_{\mathcal{C}} d$

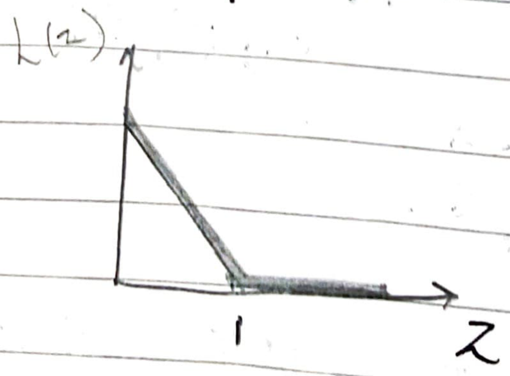
Now using conic duality the support function of $\mathbb{X}$ can be expressed as

$$S_{\mathbb{X}}(z) = \sup_x \{\langle z, x\rangle : \mathcal{C}x \preceq_{\mathcal{C}} d\}$$

$$= \inf_{q \in \mathcal{C}^*} \{\langle q, d\rangle : C^T q = z\}$$

Strong duality (thus equality above) holds bec $\mathbb{X}$ is assumed to admit a Slater point.

We consider the hinge Loss for the SVM. It is a piecewise linear loss function as below $L(z) = \max(0, 1-z)$



$$L(z) = \begin{cases} 0 & z \geq 1 \Rightarrow a_j \cdot x + b_j \Rightarrow a_j = 0,\ b_j = 0 \quad (j=1) \\ 1 - z & z \leq 1 \Rightarrow a_j \cdot x + b_j \Rightarrow a_j = -1,\ b_j = 1 \quad (j=0) \end{cases}$$

Now using constraints of the general formulation.

- $S \times (a_j \hat{y}_i w - p_{ij}^{+}) + b_j + \langle p_{ij}^{+}, \hat{x}_i \rangle \leq S_i$

for $a_j = -1$, $b_j = 1$

$j=0 \Rightarrow$ $$S \times (-\hat{y}_i w - p_{ij}^{+}) + 1 + \langle p_{ij}^{+}, \hat{x}_i \rangle \leq S_i \quad \text{-(i)}$$

for $a_j = 0$, $b_j = 0$

$j=1 \Rightarrow$ $$S \times (-p_{ij}^{+}) + \langle p_{ij}^{+}, \hat{x}_i \rangle \leq S_i \quad \text{-(ii)}$$

- $S \times (-a_j \hat{y}_i w - p_{ij}^{-}) + b_j + \langle p_{ij}^{-}, \hat{x}_i \rangle - K\lambda \leq S_i$

- for $a_j = -1$, $b_j = 1$

$j=0 \Rightarrow$ $$S \times (\hat{y}_i w - p_{ij}^{-}) + 1 + \langle p_{ij}^{-}, \hat{x}_i \rangle - K\lambda \leq S_i \quad \text{-(iii)}$$

for $a_j = 0$, $b_j = 0$

$j=1$ $$S \times (-p_{ij}^{-}) + 0 + \langle p_{ij}^{-}, \hat{x}_i \rangle - K\lambda \leq S_i \quad \text{- (iv)}$$

Using the defininion of conic dualing and support function X

We have from -(i)

$$S_X(-\hat{y}_i w - p_{ij}^+) + 1 + \langle p_{ij}^+, \hat{x}_i \rangle \leq s_i$$

$$\Downarrow$$

$$\left[\inf_{q \in \mathcal{C}^*} \{ \langle q_{ij}^+, d \rangle : C^T q_{ij}^+ = -\hat{y}_i w - p_{ij}^+ \} \right.$$
$$\left. \Rightarrow p_{ij}^+ = -\hat{y}_i w - C^T q_{ij}^+ \right]$$

Hence.

The constraint formulation of -i becomes.

$$S_X(-\hat{y}_i w - p_{ij}^+) + 1 + \langle p_{ij}^+, \hat{x}_i \rangle \leq s_i$$

$$\Rightarrow \inf_{q \in \mathcal{C}^*} \langle q_{ij}^+, d \rangle + 1 + \langle p_{ij}^+, \hat{x}_i \rangle \leq s_i$$

We make use of the $p_{ij}^+ = -\hat{y}_i w - C^T q_{ij}^+$ from above

$$\Rightarrow \inf_{q \in \mathcal{C}^*} \langle q_{ij}^+, d \rangle + 1 + \langle -C^T q_{ij}^+ - \hat{y}_i w, \hat{x}_i \rangle \leq s_i$$

if the above holds for some q_{ij}^+ then it must be for $\inf_{q \in \mathcal{C}^*}$ $\therefore$ we consider $\exists\, q_{ij}^+$ as a decision variable and remove the inf req here

Page

$\Rightarrow$ $\langle q_{ij}^{+}, d \rangle + 1 + \langle -C^{T} q_{ij}^{+} - \hat{y}_i w, \hat{x}_i \rangle \leq s_i$

Rearranging the terms we have

$$\boxed{1 - \hat{y}_i w^T x_i + \langle q_{ij}^{+}, (d - C^T x_i) \rangle \leq s_i}$$

the genera formulation has the constraint

$\| p_{ij}^{+} \|_{*} \leq \lambda$

which becomes $\boxed{\| C^T q_{ij}^{+} + \hat{y}_i w \|_{*} \leq \lambda}$

since from the support function we have $p_{ij}^{+} = -\hat{y}_i w - C^T q_{ij}^{+}$

Similar steps follow for (iii) and we get.

$$\boxed{1 + \hat{y}_i w^T x_i + \langle q_{ij}^{-}, d - C x_i \rangle - K\lambda \leq s_i}$$

and $\boxed{\| C^T q_{ij}^{-} - \hat{y}_i w \|_{*} \leq \lambda}$

Since q_{ij}^{+}, q_{ij}^{-} come from the support function

and are introdued as decision variable they must lie in the dual cone $\boxed{q_{ij}^{+}, q_{ij}^{-} \in C^{*}}$

Constraints (ii), (iv) for $j=1$ dont provide any constraint.

Since for $j=0$ the constraints are sensible we drop the subscript ℓ_{ij}^{+} and write the variable as p_i^{+}

Hence the final constraints and formulation is

$$\inf_{w,\lambda,s,p_i^{+},p_i^{-}} \lambda\varepsilon + (1/N)\sum_{i=1}^{N} s_i$$

s-t

$$1 - \hat{y}_i\langle w, \hat{x}_i\rangle + \langle p_i^{+}, d - C\hat{x}_i\rangle \le s_i \qquad i\in[N]$$

$$1 + y_i\langle w, \hat{x}_i\rangle + \langle p_i^{-}, d - C\hat{x}_i\rangle - k\lambda \le s_i \qquad i\in[N]$$

$$\| C^{T} p_i^{+} + \hat{y}_i w \|_{*} \le \lambda$$

$$\| C^{T} p_i^{-} - \hat{y}_i w \|_{*} \le \lambda$$

$$s_i \ge 0$$

$$p_i^{+}, p_i^{-} \in \mathcal{C}^{*}$$

Hence Proved $\square$.

- Formulations : Without Support

1.) Distributionally Robust SVM without Support (DRSVM_without Support).

$$\inf_{\substack{w,\lambda \\ s_i, p_i^+, p_i^-}} \quad \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i$$

$$\text{s.t.} \quad 1 - \hat{y}_i(\langle w, \hat{x}_i \rangle) \leq s_i \qquad i \in [N]$$

$$1 + \hat{y}_i(\langle w, \hat{x}_i \rangle) - K\lambda \leq s_i \qquad i \in [N]$$

$$s_i \geq 0 \qquad i \in [N]$$

$$\|w\|_* \leq \lambda$$

2.) Regularised SVM without Support (Not DR)
(RegularisedSVM_without Support)

$$\min_{w,\lambda} \quad \lambda \cdot \varepsilon + \frac{1}{N}\sum s_i$$

$$\text{s.t.} \quad 1 - y_i w^T x_i \leq s_i \qquad i \in [N]$$

$$\|w\|_* \leq \lambda$$

$$s_i \geq 0$$

NOTE: When K tends to ∞ in DRSVM the $-K\lambda \to -\infty$ $\therefore$ $(\to -\infty) \leq s_i$ constraint is trivially true and thus dropped which results in the exact formulation of Regularised SVM (non DR).