* $\mathbb{B}^d = \{ x \in \mathbb{R}^d \mid \|x\| \leq 1 \}$ et $\mathbb{B}^d \times \{\pm 1\}$

* $\Pr_{(x,y)\sim D}(y = 1 \mid x(j) > 0 \text{ pour certains } j \in K) = 1$ k étant un subset de règles (1)

* $\Pr_{(x,y)\sim D}(y\langle w, x\rangle \geq 1 \mid x(j) \leq 0 \text{ pour tout } j \in K) = 1$ (2)

  avec $w \in \mathbb{R}^d \mid \|w\|_2 \leq B$: Distributions (k, B)-réalisable

* Notation : $x(j)$ est le j^{th} feature de x

  $\mathbb{B}^{d,p}_B = \{ x \in \mathbb{R}^d \mid \|x\|_p \leq B \}$

* Redéfinisse ElasticNet comme ceci : $\{ w : \|w\|_1 \leq B_1 \text{ et } \|w\|_2^2 \leq B_2^2 \}$

* Huber-Norm ball : $\{ w_a + w_b : \|w_a\|_1 \leq B_1 \text{ et } \|w_b\|_2^2 \leq B_2 \}$

  ↳ Alice.

* $m_H(\varepsilon, \delta) = \min_A m_A(\varepsilon, \delta)$: $\qquad \|x\|_0 = |\{ j \mid x(j) \neq 0 \}|$

  D est réalisable si $l_{ramp}(\mathcal{H}, D) = 0$.

  D est n-réalisable si $l_{ramp}(\mathcal{H}, D) \leq n$

* Formalisme

  D est (k, B)-réalisable si on respecte (1) et (2) :

  w_a et w_b : $\Pr_{(x,y)\sim D}(y\langle w_a + w_b, x\rangle \geq 1) = 1$

  avec $\|w_a\|_2^2 \leq B^2$ et $\|w_b\|_0 \leq k$

Classe d'hypothèses découlant : l_0/l_2 et aussi marche pour l_0/l_1

$\mathcal{H}_{2,0} = \{ x \in \mathbb{B}^{d,2}_1 \longrightarrow \langle w_a + w_b, x\rangle \mid \|w_a\|_2^2 \leq B^2, \|w_b\|_0 \leq k \}$

* Greedy Algorithm

  Une règle ici c'est $\forall j \in [d]$ si $x_i(j) = 1 \Rightarrow y_i = 1$

  On dit j "couvre" l'example (x, y).

(1) Initialiser (2) Chercher les règles couvrant au moins $\frac{m}{100k(B+1)}$ examples (3) si échoue à 2, apprendre 1e regression linéaire