Regression linéaire

$$y(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_D x_D$$

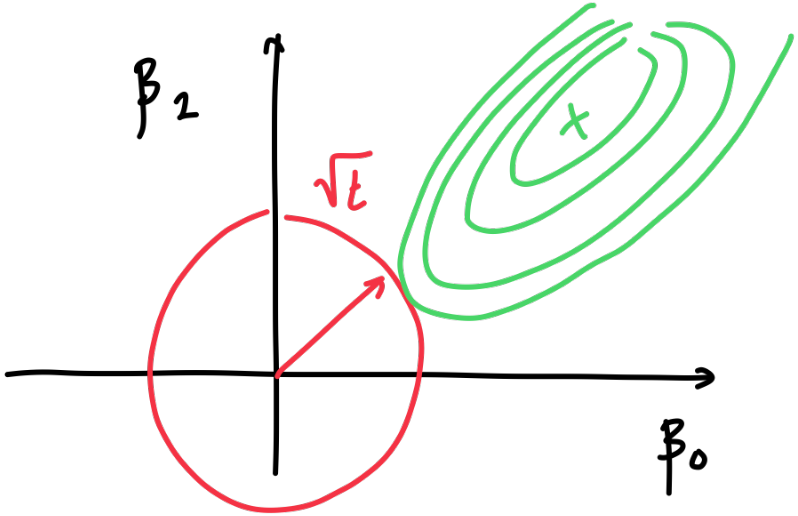
$\rightarrow$ Corrélation entre caractéristiques $\Rightarrow$ $\beta_j >>>$

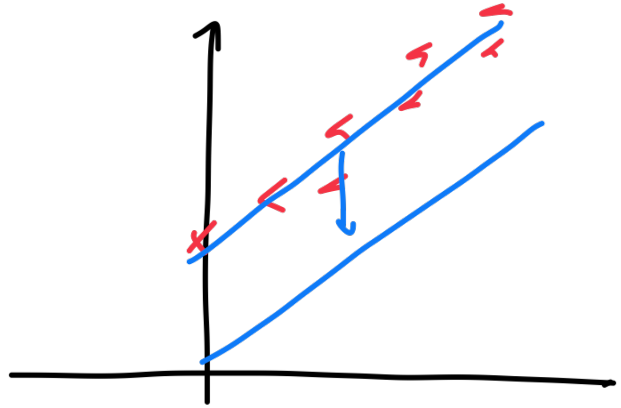
$\hookrightarrow$ Solution # 1 : Selection de caractéristiques

Solution # 2 : Contrôle de l'amplitude de β_j

Penalité RIDGE

$$\ell(\beta) = \underbrace{\frac{1}{N} \sum_{i=1}^{N} \left(t^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \dots + \beta_D x_D^{(i)} \right) \right)^2}_{\text{Fidelité aux données}} + \underbrace{\lambda \sum_{j=1}^{D} |\beta_j|^2}_{\text{Penalité sur la complexité}}$$





Pour trouver le minimum de la formulation Ridge on commence par centrer les donnée

$$t^{(i)} \leftarrow t^{(i)} - \frac{1}{N} \sum_{i=1}^{N} t^{(i)}$$

$$x_j^{(i)} \leftarrow x_j^{(i)} - \frac{1}{N} \sum_{i=1}^{N} x_j^{(i)} \qquad \longrightarrow \qquad \beta_0 = 0$$

$$\ell(\beta) = \frac{1}{N} \sum_{i=1}^{N} \left(t^{(i)} - \left(\beta_1 x_1^{(i)} + \dots + \beta_D x_D^{(i)} \right) \right)^2 + \lambda \sum_{j=1}^{D} |\beta_j|^2$$

$\rightarrow$ Seit $\vec{\beta} = (\beta_1, \beta_2, \dots, \beta_D)$ $\qquad \vec{t} = [\, t^{(2)} \dots \; t^{(N)} \,]^T$

$$\underline{\underline{X}} = \begin{bmatrix} x_1^{(2)} & \dots\dots & x_D^{(2)} \\ \vdots & & \\ \vdots & & \\ x_1^{(N)} & \dots\dots & x_D^{(N)} \end{bmatrix}$$

$$\vec{\varepsilon} = \vec{t} - \underline{\underline{X}}\,\beta$$

vecteur des résidus

$$\ell(\beta) = \frac{1}{N}\, \varepsilon^T \varepsilon + \lambda\, \beta^T \beta = \frac{1}{N} \left(\vec{t} - \underline{\underline{X}}\vec{\beta} \right)^T \left(\vec{t} - \underline{\underline{X}}\vec{\beta} \right) + \lambda\, \vec{\beta}^T \vec{\beta}$$

$$\ell(\beta) = \frac{1}{N}\left[\vec{t}^T\vec{t} - \beta^T \underline{\underline{X}}^T \vec{t} - \vec{t}^T \underline{\underline{X}} \beta + \vec{\beta}^T \underline{\underline{X}}^T \underline{\underline{X}} \beta + \lambda \vec{\beta}^T \vec{\beta}\right]$$

$$\frac{\partial \ell}{\partial \beta} = 0 - \cancel{2} \underline{\underline{X}}^T \vec{t} + \cancel{2} \underline{\underline{X}}^T \underline{\underline{X}} \vec{\beta} + \cancel{2} \lambda \vec{\beta} = 0$$

$$(\underline{\underline{X}}^T \underline{\underline{X}} + \lambda I) \vec{\beta} = \underline{\underline{X}}^T \vec{t}$$

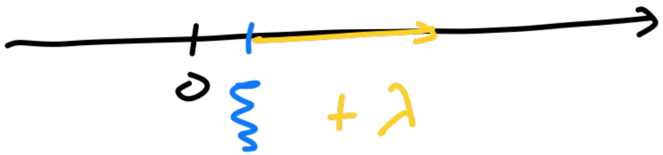
→ caractéristiques corrélées / colonnes linéairement dépendantes

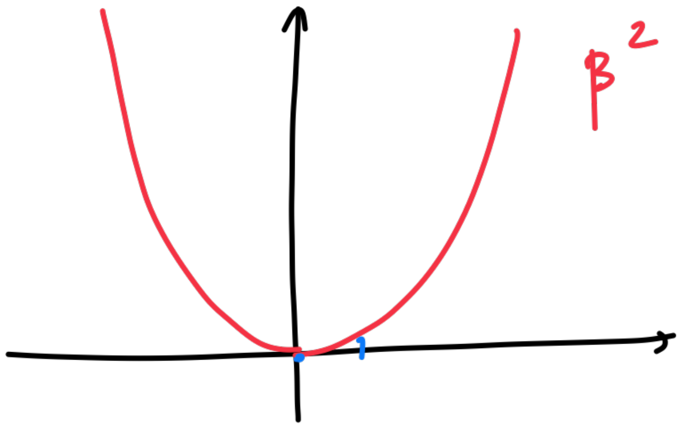
$\exists\, v$ t.q $\underline{\underline{X}}^T \underline{\underline{X}}\, v = 0$ ($\exists\, \lambda = 0$ t.q $\underline{\underline{X}}^T \underline{\underline{X}}\, v = \lambda v$)

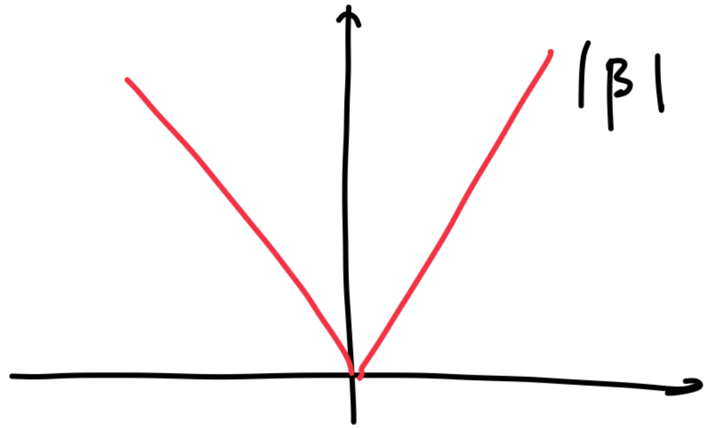
Soit (x, ξ) vecteur propre / valeur propre de $\underline{\underline{X}}^T \underline{\underline{X}}$

$\forall\,(v, \lambda)$ on a $\quad (\underline{\underline{X}}^T \underline{\underline{X}} + \lambda I)\, v = \xi v + \lambda v = (\xi + \lambda)\, v$

Quand on ajoute λI à $\underline{\underline{X}}^T \underline{\underline{X}}$ on translate les valeurs propres de $\underline{\underline{X}}^T \underline{\underline{X}}$ de λ







→ Régularisation LASSO

$$\ell(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left(t^{(i)} - (\beta_0 + \beta_1 x_1^{(i)} + \dots + \beta_D x_D^{(i)})\right)^2 + \lambda \sum_{j=1}^{D}|\beta_j|$$

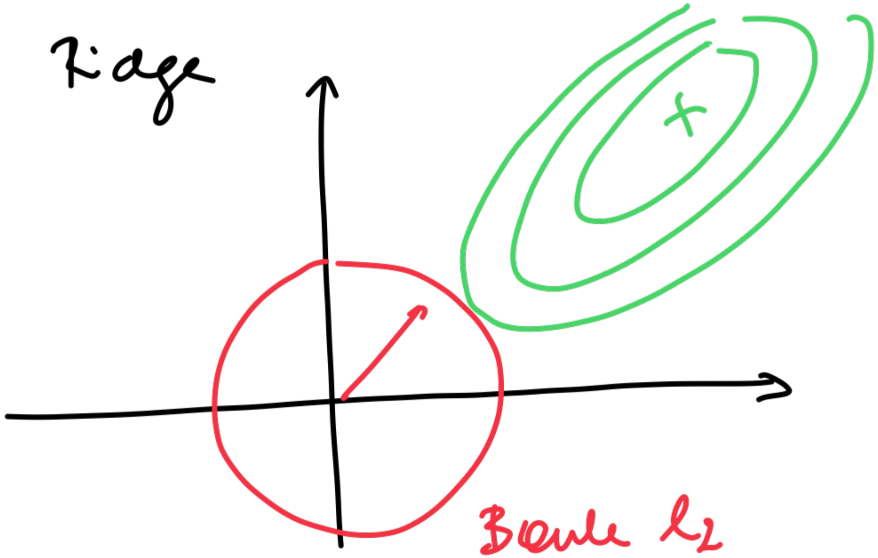
les pénalités Ridge et LASSO sont des cas particuliers de normes ℓ_p

$$\|\vec{\beta}\|_p = \left(\sum_{j=1}^{D} |\beta_j|^p\right)^{1/p}$$

Ridge ⇒ norme ℓ_2 $\|\vec{\beta}\|_2^2$

LASSO ⇒ norme ℓ_1 $\|\vec{\beta}\|_1$

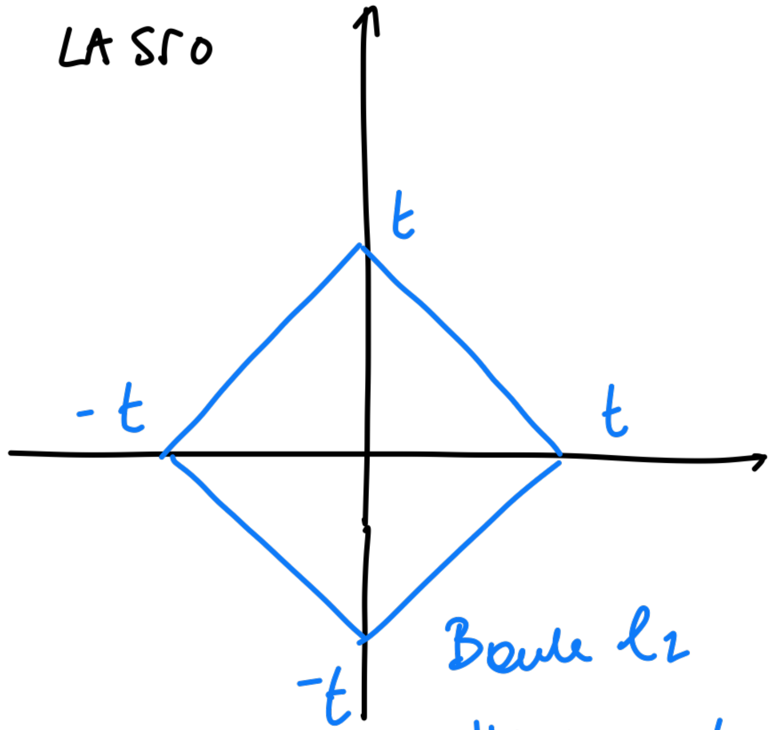
Ridge



Boule ℓ_2

$\|\beta\|_{\ell_2}^2 \leq t$

LASSO



Boule ℓ_2

$\|\beta\|_{\ell_1} \leq t$

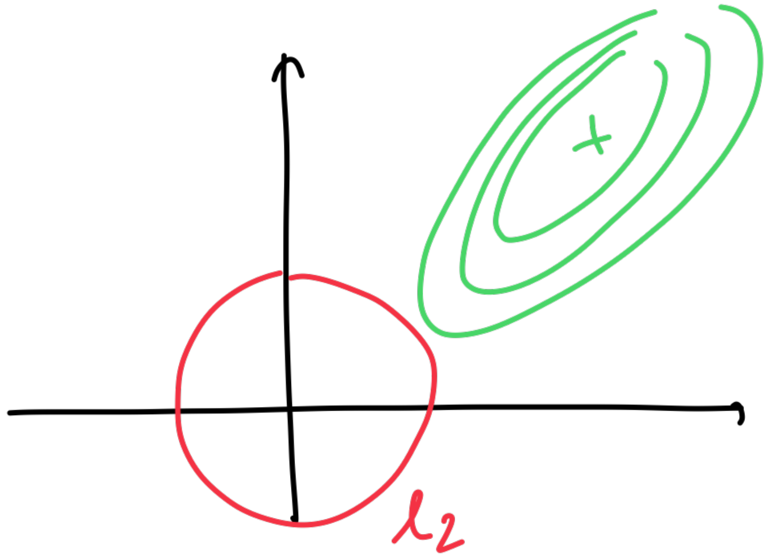
Formulation Contrainte

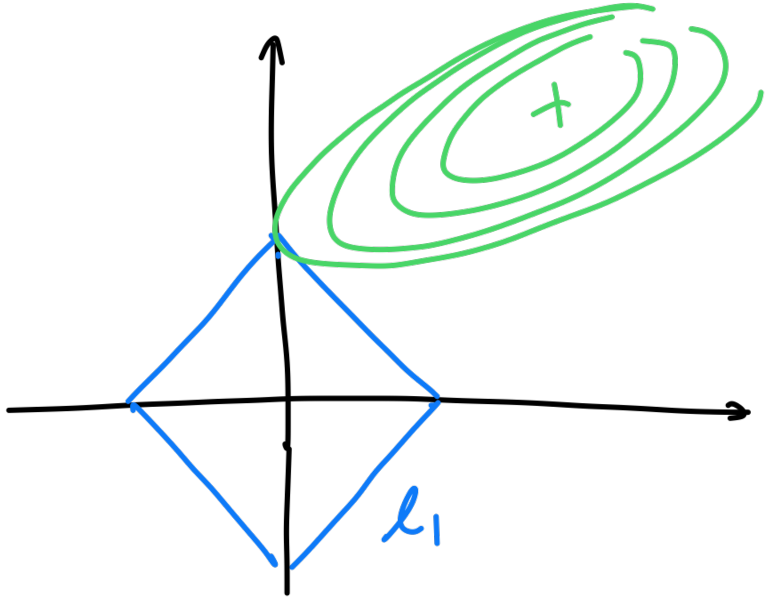
$$\min \frac{1}{N} \sum_{i=1}^{N} \left(t^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \dots + \beta_D x_D^{(i)} \right) \right)^2$$

$$\text{s.t} \quad \sum_{j=1}^{D} |\beta_j| \leq t$$

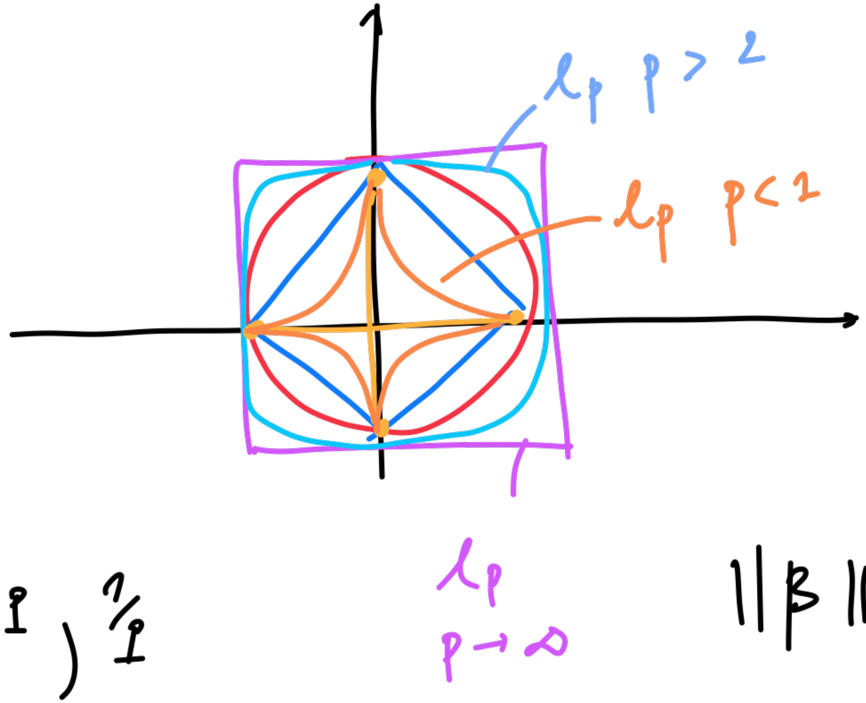
"Boule ℓ_1"

$\|\vec{\beta}\|_{\ell_1} \leq t$





Prenons $t = 1$



$$\|\beta\|_\infty \leq t$$
$$\Updownarrow$$
$$\max_j |\beta_j| \leq t$$

$$\|\beta\|_{\ell_p} = \left(\sum |\beta_j|^p\right)^{1/p}$$

$$\|\beta\|_\infty = \max_j |\beta_j|$$