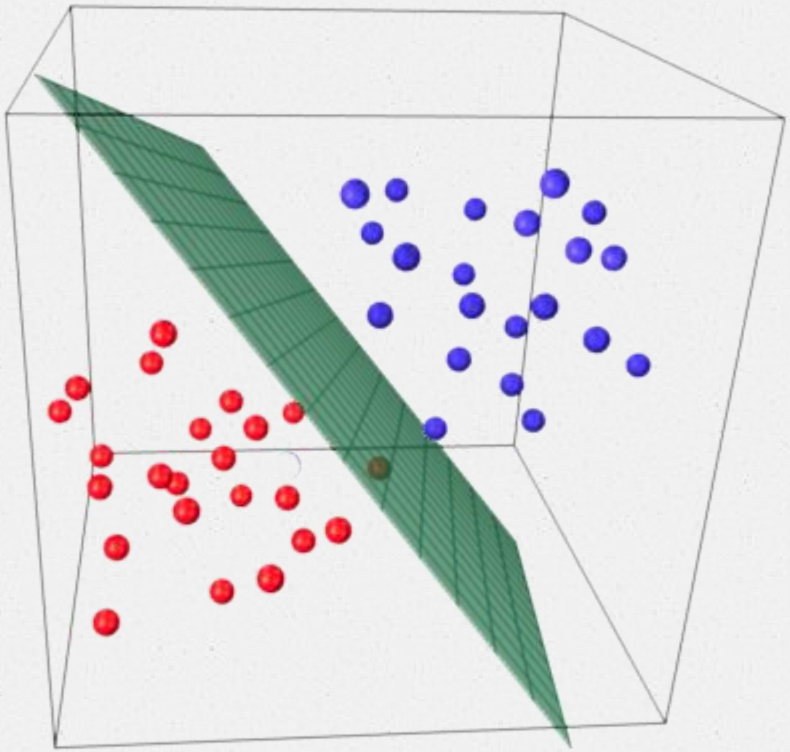
# Régression logistique en grande dimension



# Plan

- Définition & but: Régression logistique
- Simulation des données
- Calcul des gradients
- Descente de gradient par coordonnée
- Descente de gradient par coordonnée avec une pénalisation

# Définition

- Modèle sur la loi Yi sachant Xi:

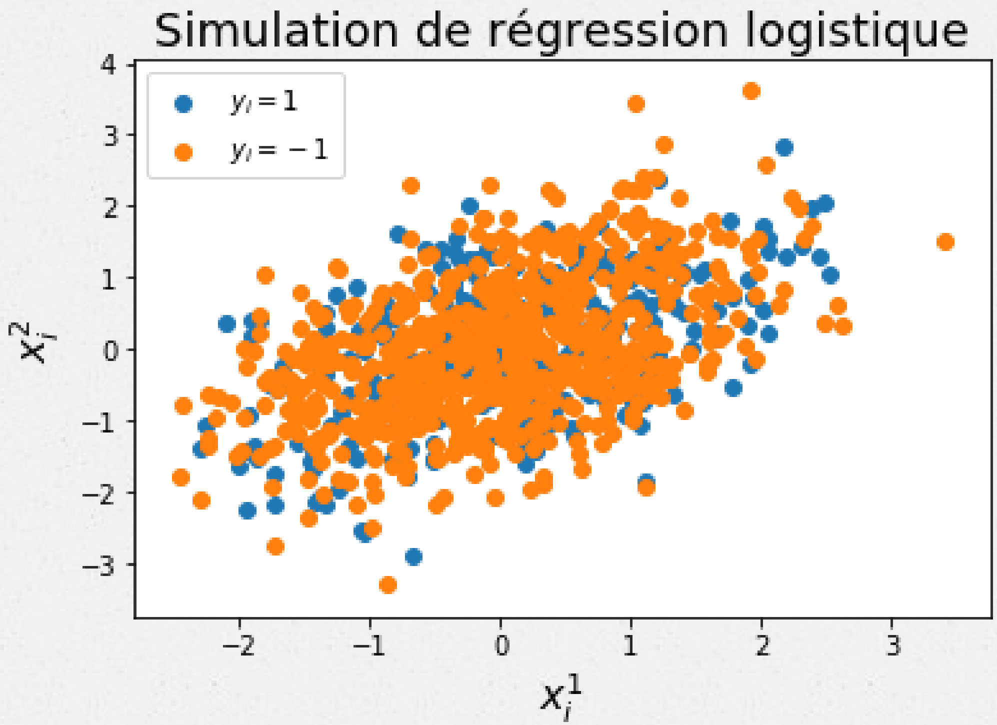
$$\mathbb{P}(y_i = 1 \mid x_i) = \sigma(x_i^\top w + b) \quad \text{avec} \quad \sigma(z) = \frac{1}{1+e^{-z}}$$

- But: estimer w et b en maximisant la vraisemblance conditionnelle Yi sachant Xi

$$f(w,b) = \frac{1}{n}\sum_{i=1}^{n} \ell(y_i, x_i^\top w + b)$$

où $\ell(y,y') = \log(1+e^{-yy'})$ : la perte logistique

# Simulation des données




- n = 1000
- d = 2
- Xi gaussiennes de la loi $N(0, \Sigma)$
- b = -1
- $w_j = (-1)^{j-1} e^{-j/10}$

# Calcul des gradients

- La dérivée première de $f(w)$ par rapport à $w_j$

$$\frac{\partial f(w)}{\partial w_j} = \frac{1}{n}\sum_{i=1}^{n}\frac{-y_i x_i^j e^{-y_i x_i^\top w_j}}{1+e^{-y_i x_i^\top w_j}}$$

$$= \frac{1}{n}\sum_{i=1}^{n} y_i x_i^j \left(\frac{-e^{-y_i x_i^\top w_j}}{1+e^{-y_i x_i^\top w_j}}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} y_i x_i^j \left(1 - \frac{1}{1+e^{-y_i x_i^\top w_j}}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} y_i x_i^j (1 - \sigma(y_i x_i^\top w_j))$$

# Descente de gradients par coordonnée

- Permutation
- Descente de gradient pour la coordonnée:

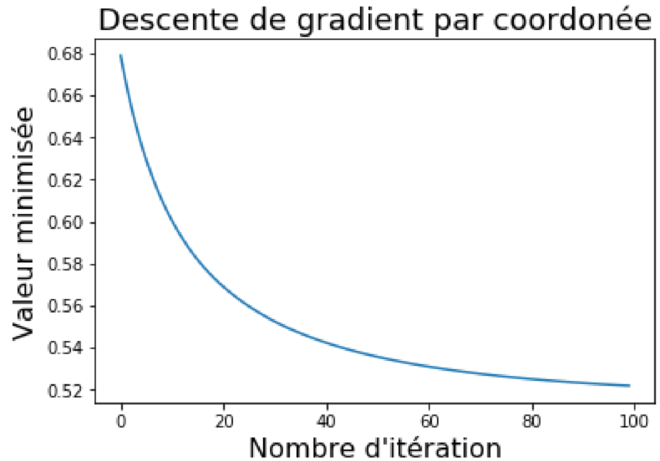
$$w_j = w_j - \eta_j \frac{\partial f}{\partial w_j}(w)$$

- Pas de descente:

$$\frac{\partial f^2}{\partial w_j^2}(w) \leq \frac{\|x^j\|_2^2}{4n}$$

# Descente de gradients par coordonnée

- *<u>Application de l'algorithme</u>*




```
Fonction minimisée 0.5218959782434925
Estimation des coefficients: [-0.75464262  0.78532741 -0.56742712  0.46433992 -0.42892956  0.45837016
 -0.41484871  0.16703482 -0.11996869  0.22916815 -0.1741995  -0.06556425
  0.10236105 -0.06260438  0.0469528  -0.05263898 -0.02836842 -0.04624102
  0.09878094 -0.13044413  0.14652444 -0.01121223  0.09414709 -0.08827123
 -0.02848357 -0.0948177   0.05506747  0.15130148 -0.14946212 -0.01978351
  0.03050244  0.00580419  0.07838211  0.04819671 -0.04697173  0.00224675
 -0.00083915 -0.09416361 -0.03607725  0.00187564  0.09980956 -0.07353569
 -0.07287572  0.05340519 -0.1051958   0.13670072 -0.0479342   0.0993725
  0.0080196  -0.06584028  0.        ]
```

# Descente de gradients par coordonnée avec pénalisation

- Technique fondamentale: utilisation d'une pénalisation

→Pénalisation Lasso:

$$\frac{1}{n}\sum_{i=1}^{n}\ell(y_i, x_i^\top w + b) + \lambda\|w\|_1$$

Avec $\lambda$ niveau de pénalisation et

$$\|w\|_1 = \sum_{j=1}^{d}|w_j|.$$

# Descente de gradients par coordonnée avec pénalisation

- Nouvelle descente de gradient:

$$w_j = s_{\lambda\eta_j}\left(w_j - \eta_j \frac{\partial f}{\partial w_j}(w)\right)$$

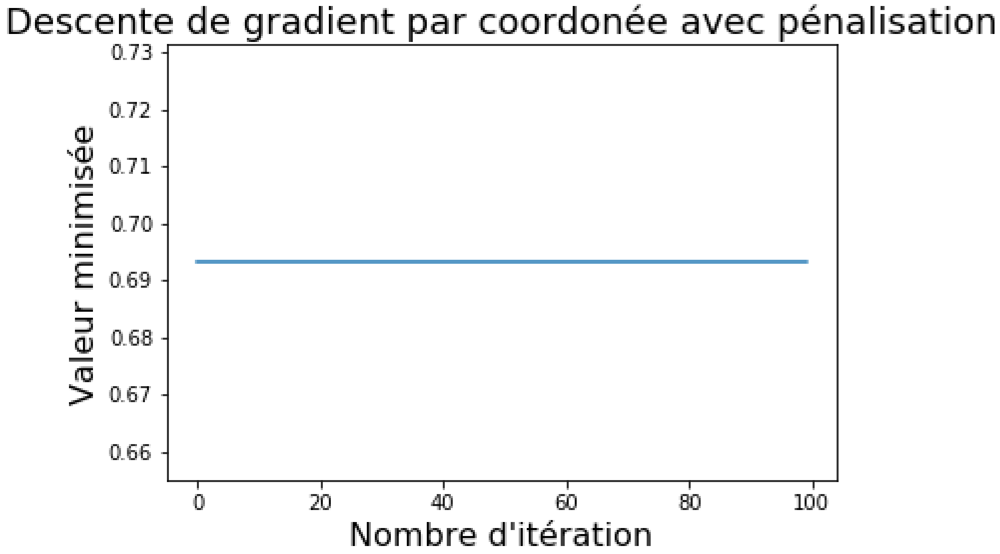
avec la fonction de *seuillage doux:*

$$s_t(z) = \mathrm{sign}(z)(|z| - t)_+$$

- Application de l'algorithme avec plusieurs niveaux de pénalisation différents

# Descente de gradients par coordonnée avec pénalisation

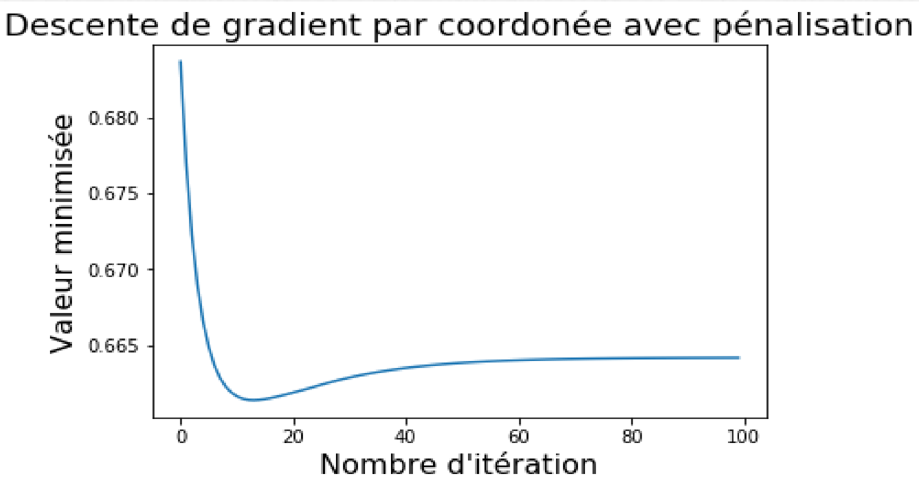
- Pour lambda = 0,5:




```
Fonction minimisée 0.6931471805599322
Estimation des coefficients: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0.]
```

# Descente de gradients par coordonnée avec pénalisation

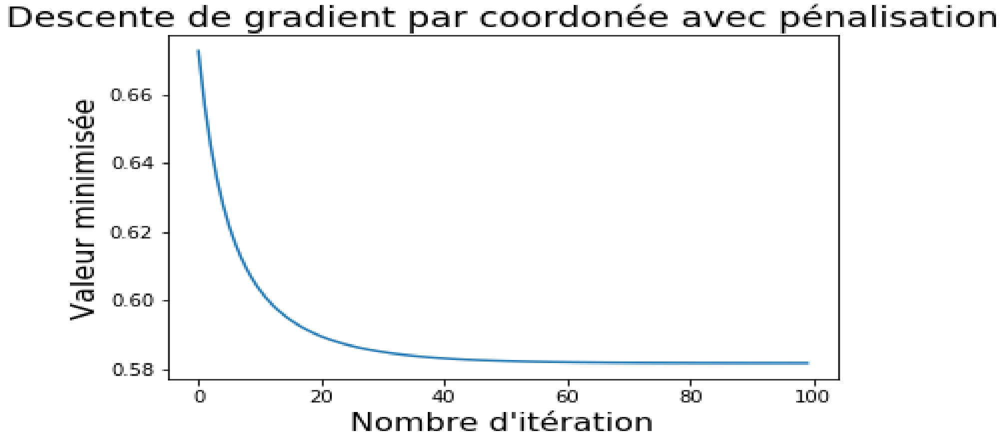
- Pour lambda = 0,05:




```
Fonction minimisée 0.5951816554165495
Estimation des coefficients: [-0.46428314  0.40318878 -0.15001747  0.11289018 -0.08967819  0.09139171
 -0.06809436  0.          0.          0.          0.          0.
  0.          0.          0.          0.          0.          0.
  0.          0.          0.          0.          0.          0.
  0.          0.          0.          0.          0.          0.
  0.          0.          0.          0.          0.          0.
  0.          0.          0.          0.          0.          0.
  0.          0.          0.          0.          0.          0.
  0.          0.          0.        ]
```

# Descente de gradients par coordonnée avec pénalisation

- Pour lambda = 0,01:




```
Fonction minimisée 0.5306567617094848
Estimation des coefficients: [-6.90770882e-01  7.24605888e-01 -5.22987424e-01  4.12968936e-01
 -3.85033898e-01  4.06033485e-01 -3.53436411e-01  1.12496162e-01
 -7.24200722e-02  1.58810662e-01 -1.31273787e-01 -2.32826153e-02
  3.51443273e-02 -3.94942466e-03  0.00000000e+00 -1.64624247e-02
 -2.09767657e-02 -2.21620215e-04  2.90799822e-02 -5.81309580e-02
  1.00937813e-01  0.00000000e+00  3.59043458e-02 -5.04795540e-02
 -1.43937834e-03 -4.79521763e-02  1.21492777e-02  1.11492045e-01
 -1.03918724e-01  0.00000000e+00  6.95332034e-03  0.00000000e+00
  6.11390632e-02  9.42681464e-03  0.00000000e+00  0.00000000e+00
  0.00000000e+00 -7.82640092e-02 -4.02964320e-03  0.00000000e+00
  4.47495819e-02 -3.67683503e-02 -3.07611991e-02  0.00000000e+00
 -4.49847623e-02  8.55899949e-02  0.00000000e+00  4.00902326e-02
  0.00000000e+00 -3.44104842e-02  0.00000000e+00]
```