Fonctions discriminantes linéaires

- Objectif
  - déterminer directement les fonctions discriminantes
  - linéaires: $g(\mathbf{x}) = w_0 + \sum_{i=1}^{d} w_i x_i = \mathbf{w}^t \mathbf{x} + w_0$
  - linéaires généralisées: $g(\mathbf{x}) = \sum_{i=1}^{\widehat{d}} a_i y_i(\mathbf{x}) = \mathbf{a}^t \mathbf{y}$
  - en minimisant le risque empirique

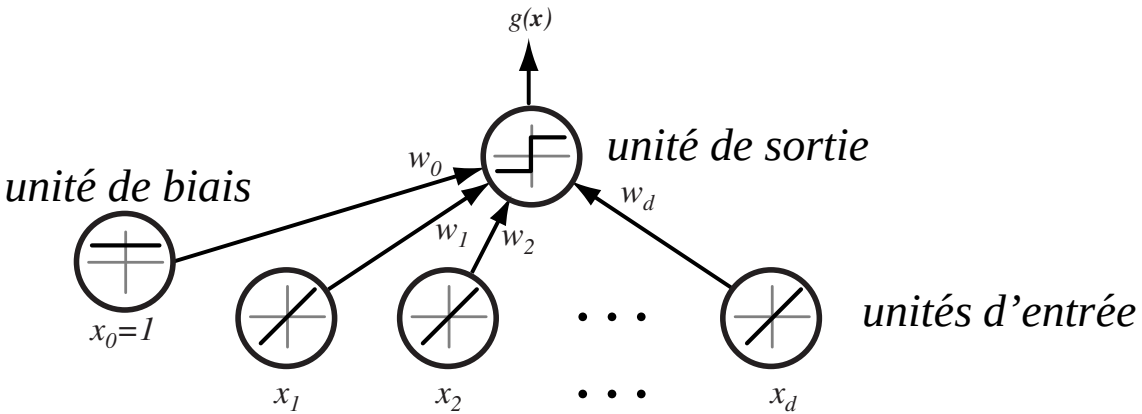
Fonctions discriminantes linéaires

- Justifications
  - parfois optimal
  - facile à calculer
  - candidates pour des classifieurs initiales
  - aborder quelques principes importants

Fonctions discriminantes linéaires

- Géométrie – deux classes
  - fonction de décision:

$$f(\mathbf{x}) = \begin{cases} C_1 & \text{si } g(\mathbf{x}) > 0, \\ C_2 & \text{si } g(\mathbf{x}) < 0 \end{cases} = \begin{cases} C_1 & \text{si } \mathbf{w}^t\mathbf{x} > -w_0, \\ C_2 & \text{si } \mathbf{w}^t\mathbf{x} < -w_0 \end{cases}$$



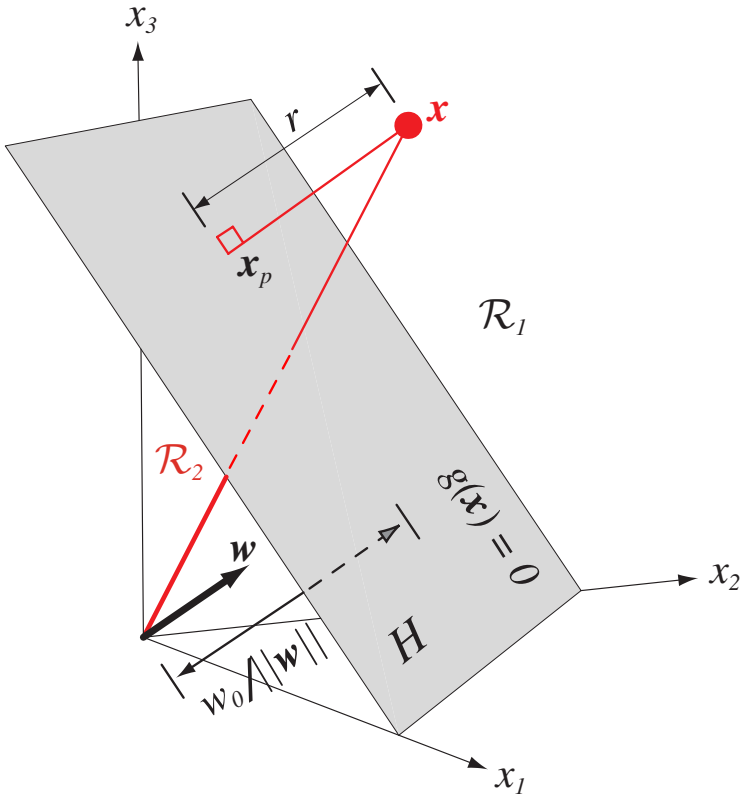
Fonctions discriminantes linéaires

- Géométrie – deux classes
  - frontière de décision H est un hyperplan:$g(\mathbf{x}) = 0$
  - $\mathbf{x}_1, \mathbf{x}_2 \in H$: $\mathbf{w}^t(\mathbf{x}_1 - \mathbf{x}_2) = 0$
  - régions de décision: R_1:coté positif, R_2:coté négatif
  - $r =$ distance algébrique de $\mathbf{x}$ et H:

$$\begin{aligned} \mathbf{x} &= \mathbf{x}_p + r\frac{\mathbf{w}}{\|\mathbf{w}\|} \\ g(\mathbf{x}) &= \mathbf{w}^t\mathbf{x} + w_0 = r\|\mathbf{w}\| \\ r &= \frac{g(\mathbf{x})}{\|\mathbf{w}\|} \end{aligned}$$

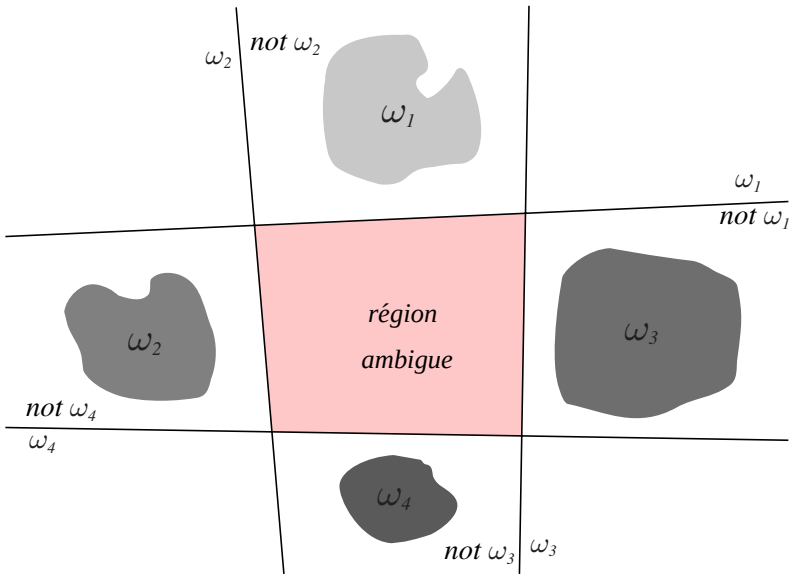
Fonctions discriminantes linéaires

- Géométrie – deux classes



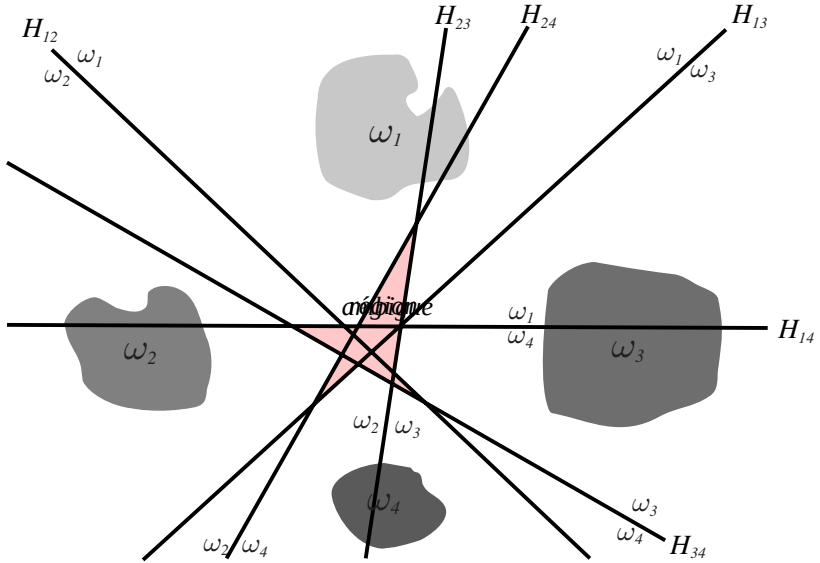
Fonctions discriminantes linéaires

- Géométrie – multiclasses
  - $C_i / \text{non} C_i$



Fonctions discriminantes linéaires

- Géométrie – multiclasses
  - $N(N-1)/2$ fonctions discriminantes

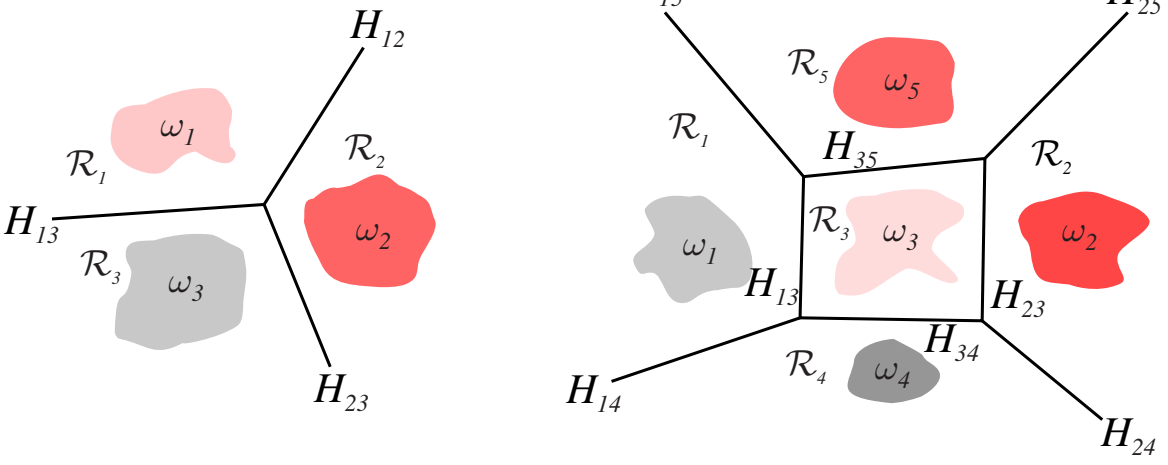


Fonctions discriminantes linéaires

- Fonctions discriminantes linéaires
  - machine linéaire: $g_j(\mathbf{x}) = \mathbf{w}_j^t \mathbf{x} + w_{j0}, \quad j = 1, \ldots, N$
  - frontières de décision $H_{i,j}$:$g_i(\mathbf{x}) = g_j(\mathbf{x})$
  - $(\mathbf{w}_i - \mathbf{w}_j)$ est orthogonal à $H_{i,j}$
  - $r(\mathbf{x}, H_{i,j}) = \dfrac{g_i(\mathbf{x}) - g_j(\mathbf{x})}{\|\mathbf{w}_i - \mathbf{w}_j\|}$

Fonctions discriminantes linéaires

- Fonctions discriminantes linéaires



Fonctions discriminantes linéaires

- Fonctions discriminantes linéaires généralisées:

$$g(\mathbf{x}) = \sum_{i=1}^{\widehat{d}} a_i y_i(\mathbf{x}) = \mathbf{a}^t \mathbf{y}$$

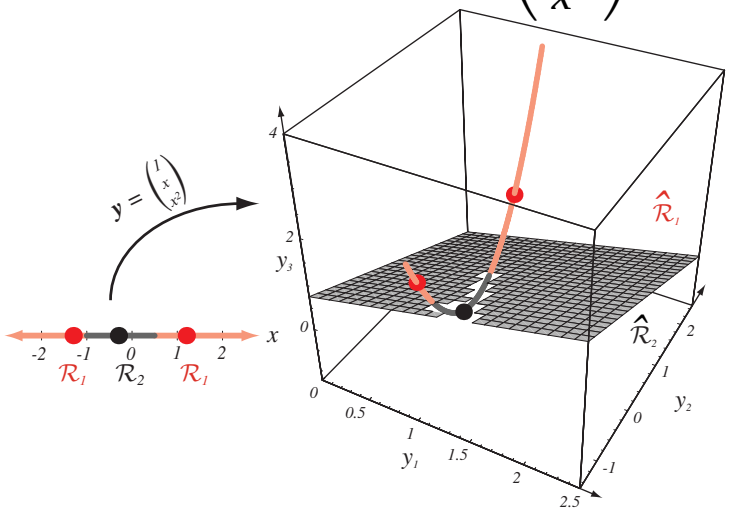
  - exemple: fonction discriminante quadratique:

$$g(\mathbf{x}) = w_0 + \sum_{i=1}^{d} w_i x_i + \sum_{i=1}^{d} \sum_{j=1}^{d} w_{ij} x_i x_j$$

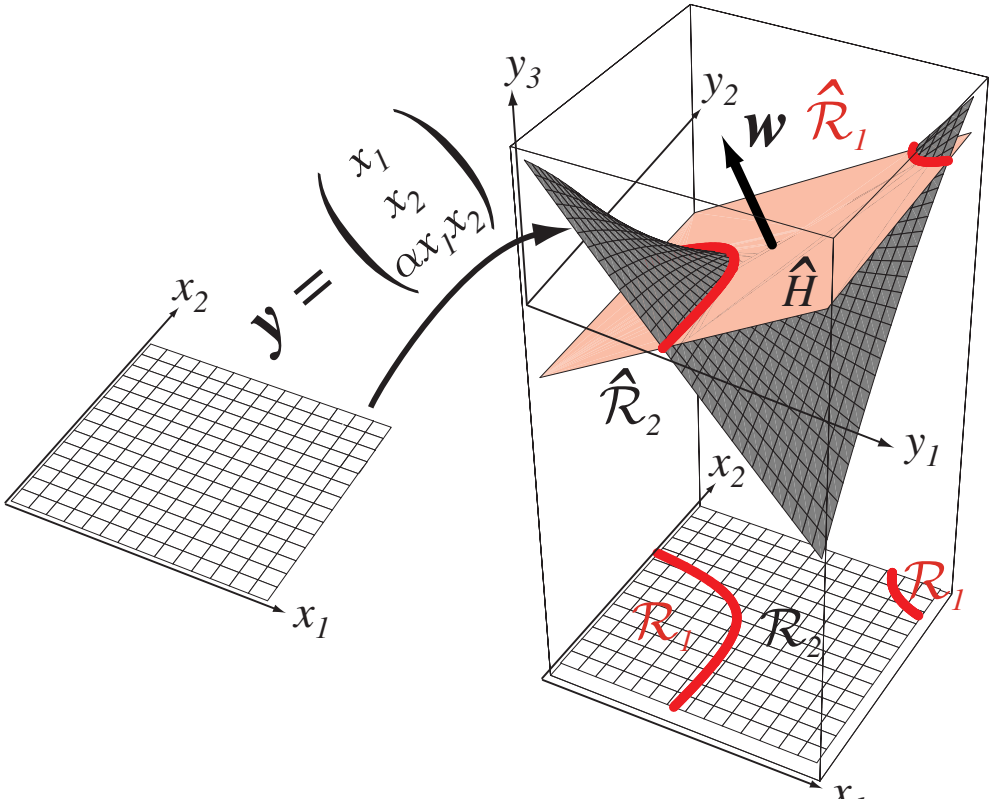
  - frontière de décision: hyperquadrique

Fonctions discriminantes linéaires

- Fonctions discriminantes linéaires généralisées
  - exemple: $g(x) = a_1 + a_2 x + a_3 x^2$, $\mathbf{y} = \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}$



- exemple: $\mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \alpha x_1 x_2 \end{pmatrix}$

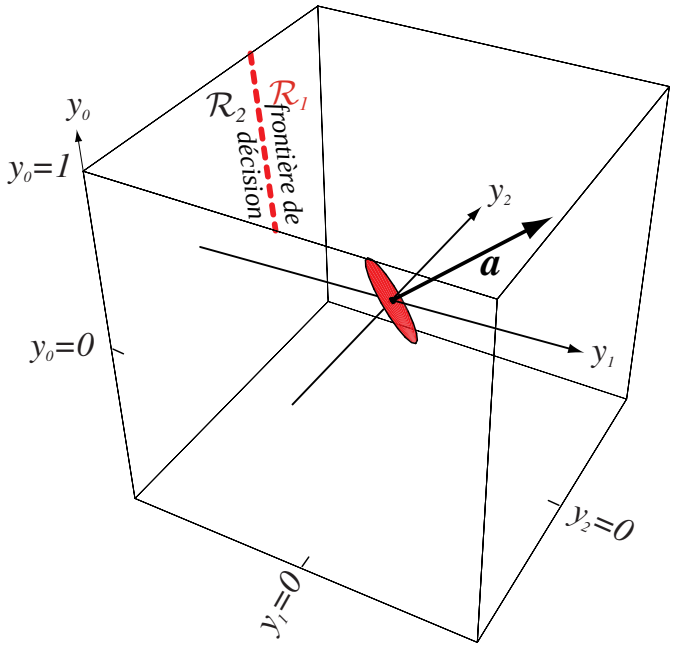


Fonctions discriminantes linéaires

- Vecteur augmenté
  - $g(\mathbf{x}) = w_0 + \sum_{i=1}^{d} w_i x_i = \sum_{i=0}^{d} w_i x_i \quad (x_0 = 1)$
  - $g(\mathbf{x}) = \sum_{i=1}^{\widehat{d}} a_i y_i,\ \widehat{d} = d+1,\ \mathbf{y} = \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{pmatrix},\ \mathbf{a} = \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{pmatrix} = \begin{pmatrix} w_0 \\ \\ \mathbf{w} \\ \\ \end{pmatrix}$

Fonctions discriminantes linéaires

- Vecteur augmenté



Fonctions discriminantes linéaires

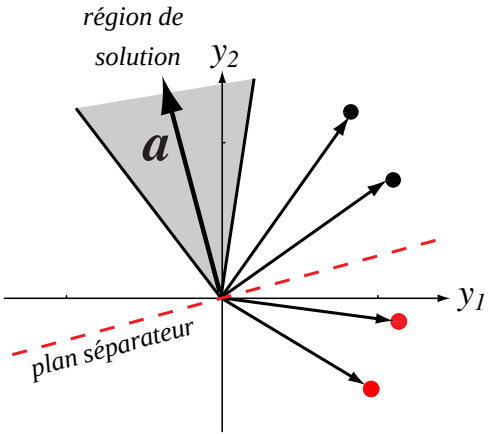
- Séparabilité linéaire
  - $D_n = \big((\mathbf{y}_1, z_1), \ldots, (\mathbf{y}_n, z_n)\big)$, $z_i = \begin{cases} 1 & \text{si } \mathbf{y}_i \text{ est classifié } C_1 \\ -1 & \text{si } \mathbf{y}_i \text{ est classifié } C_2 \end{cases}$
  - $g(\mathbf{x}) = \mathbf{a}^t\mathbf{y}$ sépare D_n sans erreur:

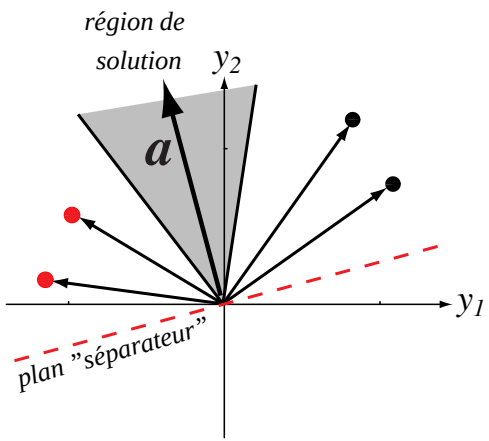
$$\mathbf{a}^t\mathbf{y}_i z_i > 0, \quad i = 1, \ldots, n$$

  - $\mathbf{a}$: vecteur séparateur, vecteur de solution

Fonctions discriminantes linéaires

- Séparabilité linéaire





Fonctions discriminantes linéaires

- Marge de séparation:

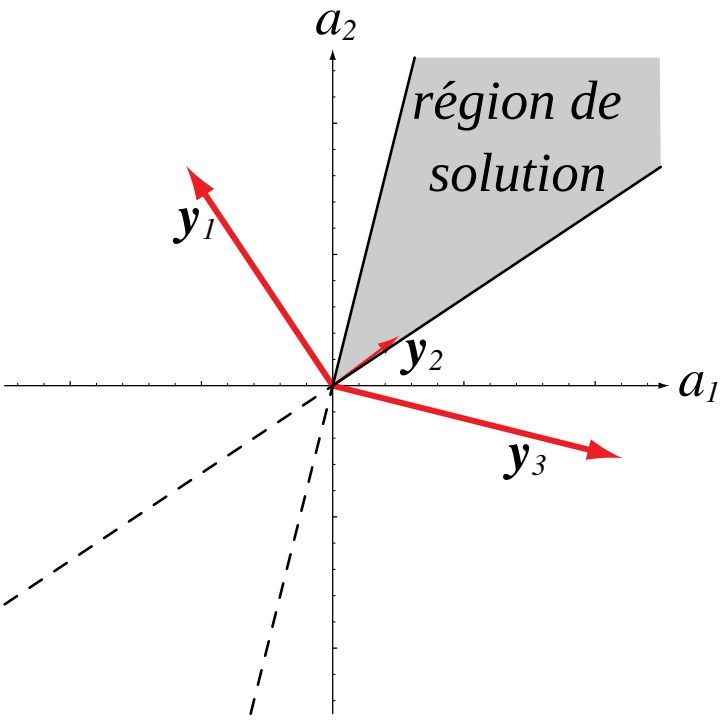
$$m_i = g(\mathbf{x}_i)z_i = \mathbf{a}^t\mathbf{y}z_i$$

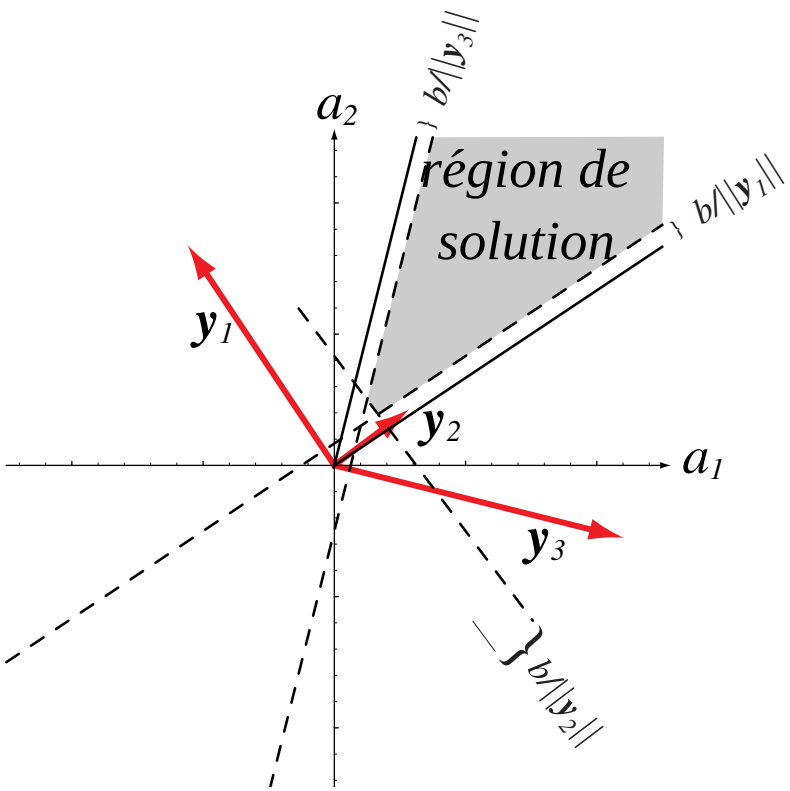
  - Séparation avec une marge b:

$$m_i = \mathbf{a}^t\mathbf{y}_i z_i > b, \quad i = 1, \ldots, n$$

Fonctions discriminantes linéaires

- Marge de séparation





Fonctions discriminantes linéaires

- Procédures de descente de gradient
  - fonction de critère: $J(\mathbf{a})$ – minimisée si $\mathbf{a}$ est une solution
  - $\mathbf{a}(k+1) = \mathbf{a}(k) - \eta(k)\nabla J(\mathbf{a}(k))$
  - $\eta(k)$: taux d'apprentissage

DESCENTEDEGRADIENTSIMPLE$(\Theta, \eta(\cdot), \mathbf{a}_0)$

1 $\mathbf{a} \leftarrow \mathbf{a}_0$
2 $k \leftarrow 0$
3 **faire**
4 $\quad k \leftarrow k+1$
5 $\quad \mathbf{a} \leftarrow \mathbf{a} - \eta(k)\nabla J(\mathbf{a})$
6 **jusqu'à** $|\eta(k)\nabla J(\mathbf{a})| < \Theta$
7 **retourner a**

Fonctions discriminantes linéaires

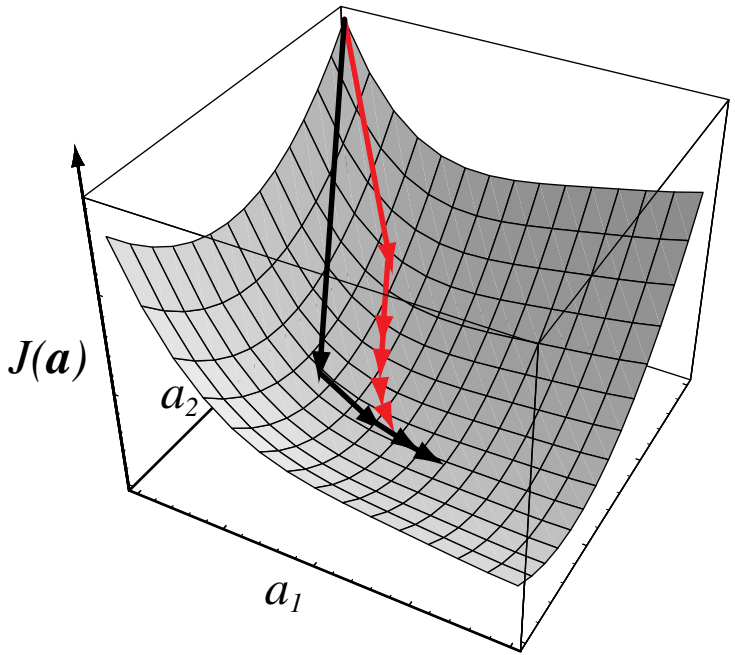
- Descente de Newton
  - $J(\mathbf{a}) \simeq J(\mathbf{a}(k)) + \nabla J^t(\mathbf{a} - \mathbf{a}(k)) + \frac{1}{2}(\mathbf{a} - \mathbf{a}(k))^t \mathbf{H}(\mathbf{a} - \mathbf{a}(k))$
  - matrice hessienne: $\mathbf{H}_{ij} = \frac{\delta^2 J}{\delta a_i \delta a_j}$
  - $\mathbf{a}(k+1) = \mathbf{a}(k) - \mathbf{H}^{-1} \nabla J$

DESCENTEDENEWTON($\Theta, \mathbf{a}_0$)

```
1  a ← a_0
2  faire
3      a ← a − H^{-1} ∇J(a)
4  jusqu'à |H^{-1} ∇J(a)| < Θ
5  retourner a
```

Fonctions discriminantes linéaires

- Descente de Newton



Fonctions discriminantes linéaires

- Le perceptron
  - $J_p(\mathbf{a}) = \sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq 0\}} (-\mathbf{a}^t \mathbf{y}_i z_i)$
  - $\nabla J_p = \sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq 0\}} (-\mathbf{y}_i z_i)$
  - $\mathbf{a}(k+1) = \mathbf{a}(k) + \eta(k) \sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq 0\}} \mathbf{y}_i z_i$

Fonctions discriminantes linéaires

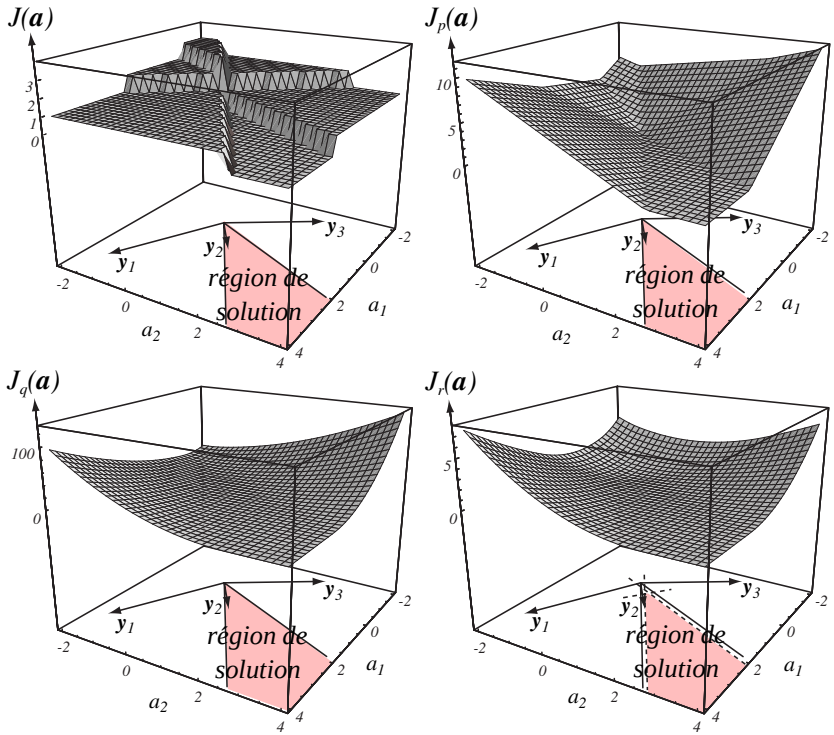
- Le perceptron

PERCEPTRONBATCH$(\Theta, \eta(\cdot), \mathbf{a}_0)$

1 $\mathbf{a} \leftarrow \mathbf{a}_0$
2 $k \leftarrow 0$
3 **faire**
4 $\quad k \leftarrow k+1$
5 $\quad \mathbf{a} \leftarrow \mathbf{a} + \eta(k) \sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq 0\}} \mathbf{y}_i z_i$
6 **jusqu'à** $\left|\eta(k) \sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq 0\}} \mathbf{y}_i z_i\right| < \Theta$
7 **retourner a**

Fonctions discriminantes linéaires

- Fonctions de critère



Fonctions discriminantes linéaires

- Le perceptron en-ligne

PERCEPTRONENLIGNE($\mathbf{a}_0$)

1 $\mathbf{a} \leftarrow \mathbf{a}_0$
2 $k \leftarrow 0$
3 **faire**
4 $\quad k \leftarrow (k+1) \mod n$
5 $\quad$ **si** $\mathbf{a}^t \mathbf{y}_k z_k \leq 0$ **alors** $\quad \triangleright$ $\mathbf{y}_k$ *mal classifié*
6 $\quad\quad \mathbf{a} \leftarrow \mathbf{a} + \mathbf{y}_k z_k$
7 **jusqu'à** $\sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq 0\}} = 0$ $\quad \triangleright$ *pas d'erreur*
8 **retourner a**

- Théorème
  - Si l'ensemble d'entraînement est linéairement séparable, l'algorithm PERCEPTRONENLIGNE se termine à une vecteur de solution après un nombre fini de corrections.

Fonctions discriminantes linéaires

- Le perceptron en-ligne, avec marge, d'incrément variable

PERCEPTRONENLIGNEMARGEVARIABLE$(\eta(\cdot), \mathbf{a}_0, b)$

1 $\mathbf{a} \leftarrow \mathbf{a}_0$
2 $k \leftarrow 0$
3 **faire**
4 $\quad k \leftarrow k+1$
5 $\quad k' \leftarrow k \mod n$
6 $\quad$ **si** $\mathbf{a}^t \mathbf{y}_{k'} z_{k'} \leq b$ **alors**
7 $\quad\quad \mathbf{a} \leftarrow \mathbf{a} + \eta(k)\mathbf{y}_{k'} z_{k'}$
8 **jusqu'à** $\sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq b\}} = 0$ ▷ *pas d'erreur par rapport à la marge b*
9 **retourner a**

Fonctions discriminantes linéaires

- Conditions de convergence
  - $\eta(k) \geq 0$
  - $\lim_{m \to \infty} \sum_{k=1}^{m} \eta(k) = \infty$
  - $\lim_{m \to \infty} \frac{\sum_{k=1}^{m} \eta^2(k)}{\left(\sum_{k=1}^{m} \eta(k)\right)^2} = 0$

Fonctions discriminantes linéaires

- Le perceptron batch d'incrément variable
  - $\mathbf{y}^{(k)} = \sum_{i=1}^{n} I_{\{\mathbf{a}^t(k)\mathbf{y}_i z_i \leq 0\}} \mathbf{y}_i z_i$

PERCEPTRONBATCHVARIABLE$(\eta(\cdot), \mathbf{a}_0)$

1 $\mathbf{a} \leftarrow \mathbf{a}_0$
2 $k \leftarrow 0$
3 **faire**
4 $\quad k \leftarrow k+1$
5 $\quad \mathbf{a} \leftarrow \mathbf{a} + \eta(k) \sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq 0\}} \mathbf{y}_i z_i$
6 **jusqu'à** $\sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq 0\}} = 0$
7 **retourner a**

Fonctions discriminantes linéaires

- Procédures de relaxation
  - $J_q(\mathbf{a}) = \sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq 0\}} (\mathbf{a}^t \mathbf{y}_i z_i)^2$
  - $J_r(\mathbf{a}) = \frac{1}{2} \sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq b\}} \frac{(\mathbf{a}^t \mathbf{y}_i z_i - b)^2}{\|\mathbf{y}_i z_i\|^2}$
  - $\nabla J_r = \sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq b\}} \frac{\mathbf{a}^t \mathbf{y}_i z_i - b}{\|\mathbf{y}_i z_i\|^2} \mathbf{y}_i z_i$
  - $\mathbf{a}(k+1) = \mathbf{a}(k) + \eta(k) \sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq b\}} \frac{b - \mathbf{a}^t \mathbf{y}_i z_i}{\|\mathbf{y}_i z_i\|^2} \mathbf{y}_i z_i$

Fonctions discriminantes linéaires

- Procédures de relaxation

RELAXATIONBATCHMARGE($\eta(\cdot), \mathbf{a}_0, b$)

1 $\mathbf{a} \leftarrow \mathbf{a}_0$
2 $k \leftarrow 0$
3 **faire**
4 $\quad k \leftarrow k+1$
5 $\quad \mathbf{a} \leftarrow \mathbf{a} + \eta(k) \sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq b\}} \frac{b - \mathbf{a}^t \mathbf{y}_i z_i}{\|\mathbf{y}_i z_i\|^2} \mathbf{y}_i z_i$
6 **jusqu'à** $\sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq b\}} = 0$
7 **retourner a**

Fonctions discriminantes linéaires

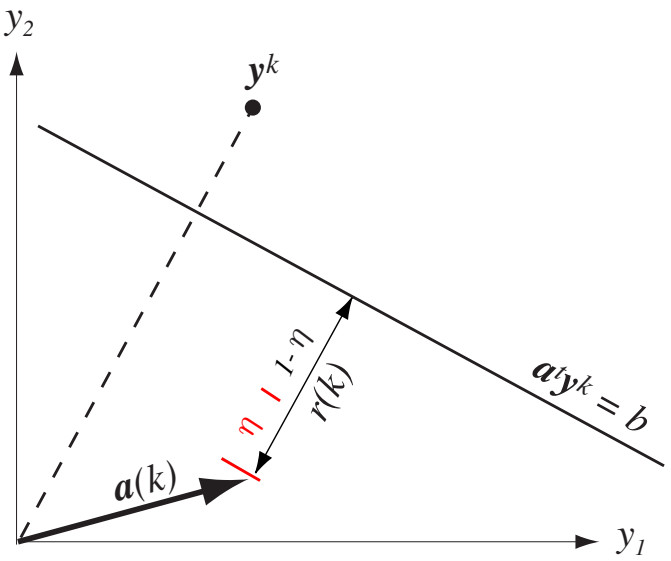
- Relaxation en-ligne

RELAXATIONENLIGNEMARGE$(\eta(\cdot), \mathbf{a}_0, b)$

1 $\mathbf{a} \leftarrow \mathbf{a}_0$
2 $k \leftarrow 0$
3 **faire**
4 $\quad k \leftarrow k+1$
5 $\quad k' \leftarrow k \mod n$
6 $\quad$ **si** $\mathbf{a}^t \mathbf{y}_{k'} z_{k'} \leq b$ **alors**
7 $\quad\quad \mathbf{a} \leftarrow \mathbf{a} + \eta(k) \frac{b - \mathbf{a}^t \mathbf{y}_{k'} z_{k'}}{\|\mathbf{y}_{k'} z_{k'}\|^2} \mathbf{y}_{k'} z_{k'}$
8 **jusqu'à** $\sum_{i=1}^{n} I_{\{\mathbf{a}^t \mathbf{y}_i z_i \leq b\}} = 0$
9 **retourner a**

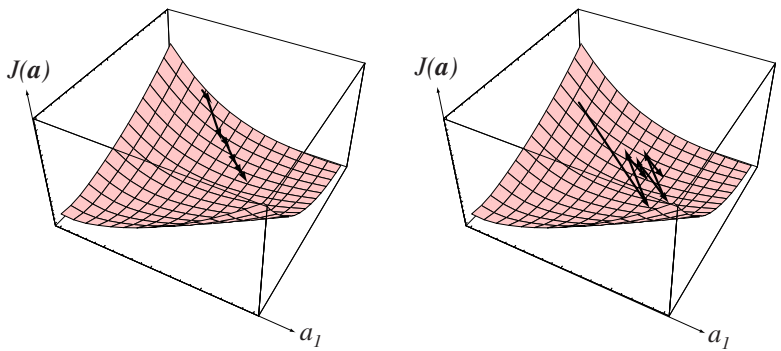
Fonctions discriminantes linéaires

- Relaxation en-ligne
  - $r(k) = \dfrac{b - \mathbf{a}^t \mathbf{y}_{k'} z_{k'}}{\|\mathbf{y}_{k'} z_{k'}\|}$



Fonctions discriminantes linéaires

- Relaxation en-ligne
  - $\eta > 1$: sur-relaxation
  - $\eta < 1$: sous-relaxation
  - condition de convergence: $0 < \eta < 2$



Fonctions discriminantes linéaires

- Comportement dans le cas non-séparable
  - procédures de correction d'erreur
  - fonctionnent bien si
    - la décision de Bayes est à peu près linéaire
    - l'erreur de Bayes est petite
  - si $2\widehat{d} > n$, la probabilité de non-séparabilité est petite

Fonctions discriminantes linéaires

- Incrément fixe
  - boucle infinie
  - engendre un procéssus d'état fini
  - moyenner les vecteurs de poids
- Incrément variable
  - converge si $\eta(k) \rightarrow 0$

Fonctions discriminantes linéaires

- L'approche d'erreur carrée (régression)
  - soit $\mathbf{b} = (z_1, \ldots, z_n)^t$
  - Idéalement on voudrait trouver $\mathbf{a}$ tel que $\mathbf{Ya} = \mathbf{b}$
  - Mais on commet des erreurs $\mathbf{e} = \mathbf{Ya} - \mathbf{b}$
  - $J_s(\mathbf{a}) = \|\mathbf{Ya} - \mathbf{b}\|^2 = \sum_{i=1}^{n} (\mathbf{a}^t \mathbf{y}_i - b_i)^2$
  - $\nabla J_s(\mathbf{a}) = \sum_{i=1}^{n} 2(\mathbf{a}^t \mathbf{y}_i - b_i)\mathbf{y}_i = 2\mathbf{Y}^t(\mathbf{Ya} - \mathbf{b})$
  - $\mathbf{Y}^t\mathbf{Ya} = \mathbf{Y}^t\mathbf{b}$
  - $\mathbf{a} = (\mathbf{Y}^t\mathbf{Y})^{-1}\mathbf{Y}^t\mathbf{b} = \mathbf{Y}^\dagger\mathbf{b}$

- $\mathbf{Y}^{\dagger} = (\mathbf{Y}^{t}\mathbf{Y})^{-1}\mathbf{Y}^{t}$: pseudoinverse de $\mathbf{Y}$

Fonctions discriminantes linéaires

- Procédure de Widrow-Hoff (LMS)
  - batch: $\mathbf{a}(k+1) = \mathbf{a}(k) + \eta(k)\mathbf{Y}^t(\mathbf{b} - \mathbf{Ya}(k))$
  - en ligne: $\mathbf{a}(k+1) = \mathbf{a}(k) + \eta(k)\mathbf{y}_{k'}(b_{k'} - \mathbf{a}^t\mathbf{y}_{k'})$

LMS$(\Theta, \eta(\cdot), \mathbf{a}_0)$

1 $\quad \mathbf{a} \leftarrow \mathbf{a}_0$
2 $\quad k \leftarrow 0$
3 $\quad$ **faire**
4 $\quad\quad k \leftarrow k+1$
5 $\quad\quad k' \leftarrow k \mod n$
6 $\quad\quad \mathbf{a} \leftarrow \mathbf{a} + \eta(k)\mathbf{y}_{k'}(b_{k'} - \mathbf{a}^t\mathbf{y}_{k'})$
7 $\quad$ **jusqu'à** $|\eta(k)\mathbf{y}_{k'}(b_{k'} - \mathbf{a}^t\mathbf{y}_{k'})| < \Theta$
8 $\quad$ **retourner a**

Fonctions discriminantes linéaires

- Procédure de Widrow-Hoff (LMS)
  - se comporte bien dans le cas non-séparable
  - ne converge pas nécessairement à un hyperplan séparateur dans les cas séparables

Fonctions discriminantes linéaires

- La machine de support vector (SVM)
  - objectif: trouver un hyperplan séparateur avec une grande marge $z_i g(\mathbf{y}_i) = z_i \mathbf{a}^t \mathbf{y}_i$
  - maximiser b: $\frac{z_i g(\mathbf{y}_i)}{\|\mathbf{a}\|} \geq b \quad i = 1, \ldots, n$