

Apprentissage de Fonctions Linéaires à Seuil

Frederic Koriche

Cours Intelligence Artificielle
Apprentissage: partie IV
Université Montpellier II, France

Frederic.Koriche@lirmm.fr

Outline

- 1 **Cadre Formel**
 - Instances
 - Fonctions Linéaires à Seuil
 - Normalisation
 - Séparabilité
- 2 Le perceptron
 - L'algorithme
 - Convergence
- 3 Winnow
 - L'algorithme
 - Convergence

Attributs et Valeurs

L'espace des observations est décrit par un ensemble fini d'attributs $\{1, \dots, n\}$ qui prennent des valeurs continues ou discrètes.

Instances et Exemples

- **Instance** : fonction x qui associe à tout attribut i une valeur x_i .
- **Exemple** : paire (x, y) où x est une instance et $y \in \{-1, 1\}$ est la classe de l'instance.

Espace

L'espace des instances est donné par X .

Attributs et Valeurs

L'espace des observations est décrit par un ensemble fini d'attributs $\{1, \dots, n\}$ qui prennent des valeurs continues ou discrètes.

Instances et Exemples

- **Instance** : fonction x qui associe à tout attribut i une valeur x_i .
- **Exemple** : paire (x, y) où x est une instance et $y \in \{-1, 1\}$ est la classe de l'instance.

Espace

L'espace des instances est donné par X .

Définition

Une fonction linéaire à seuil est une expression de la forme $(w_1, \dots, w_n, \theta)$ où :

- w_1, \dots, w_n sont des poids
- θ est une valeur appelée seuil.

Classification

Soit f une fonction linéaire à seuil et x une instance :

$$f(x) = \text{sign}(w_1x_1 + \dots + w_nx_n - \theta)$$

Définition

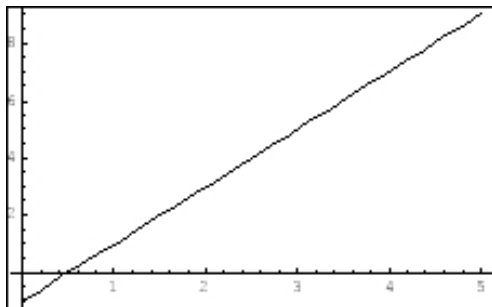
Une fonction linéaire à seuil est une expression de la forme $(w_1, \dots, w_n, \theta)$ où :

- w_1, \dots, w_n sont des poids
- θ est une valeur appelée seuil.

Classification

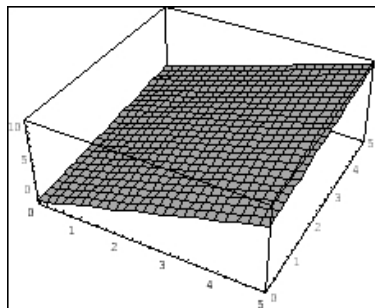
Soit f une fonction linéaire à seuil et x une instance :

$$f(x) = \text{sign}(w_1x_1 + \dots + w_nx_n - \theta)$$



Exemple

$$2x_1 - 1$$



Exemple

$$2x_1 + \frac{1}{2}x_2 - 1$$

Définition

Chaque attribut, poids et seuil prend une valeur dans $[0, 1]$.

Propriété

Toute fonction linéaire à seuil peut se transformer en une fonction normale équivalente (qui classe de la même manière).

Définition

Chaque attribut, poids et seuil prend une valeur dans $[0, 1]$.

Propriété

Toute fonction linéaire à seuil peut se transformer en une fonction normale équivalente (qui classe de la même manière).

Marge

La marge d'une fonction à seuil est donnée par :

$$\delta = \min_{x \in X} (|w_1 x_1 + \dots + w_n x_n - \theta|)$$

Séparabilité

Une fonction à seuil est linéairement séparable si $\delta > 0$.

Marge

La marge d'une fonction à seuil est donnée par :

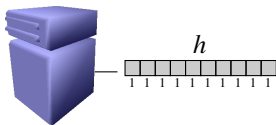
$$\delta = \min_{x \in X} (|w_1 x_1 + \dots + w_n x_n - \theta|)$$

Séparabilité

Une fonction à seuil est linéairement séparable si $\delta > 0$.

Outline

- 1 Cadre Formel
 - Instances
 - Fonctions Linéaires à Seuil
 - Normalisation
 - Séparabilité
- 2 Le perceptron
 - L'algorithme
 - Convergence
- 3 Winnow
 - L'algorithme
 - Convergence



Initialisation

- 1 Choisir les poids w_i et le seuil θ aléatoirement

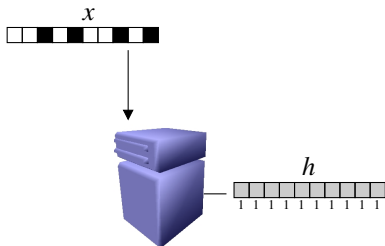
Tours

- 2 Recevoir une instance x
- 3 Envoyer la prédiction

$$h(x) = \text{sign} \left(\sum_{i=1}^n w_i x_i - \theta \right)$$

- 4 Recevoir la réponse $f(x)$. En cas d'erreur,

$$w_i \leftarrow w_i + \eta(f(x) - h(x))x_i$$
$$\theta \leftarrow \theta + \eta f(x)\theta^2$$



Initialisation

- 1 Choisir les poids w_i et le seuil θ aléatoirement

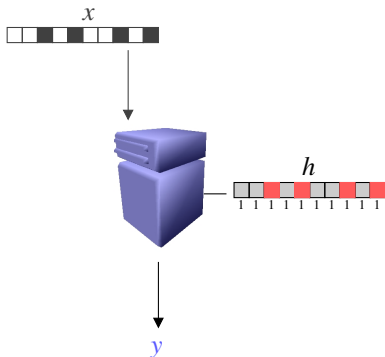
Tours

- 1 Recevoir une instance x
- 2 Envoyer la prédiction

$$h(x) = \text{sign} \left(\sum_{i=1}^n w_i x_i - \theta \right)$$

- 3 Recevoir la réponse $f(x)$. En cas d'erreur,

$$\begin{aligned} w_i &\leftarrow w_i + \eta(f(x) - h(x))x_i \\ \theta &\leftarrow \theta + \eta f(x)n^2 \end{aligned}$$



Initialisation

- 1 Choisir les poids w_i et le seuil θ aléatoirement

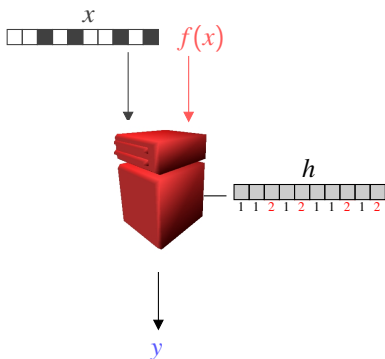
Tours

- 1 Recevoir une instance x
- 2 Envoyer la prédiction

$$h(x) = \text{sign} \left(\sum_{i=1}^n w_i x_i - \theta \right)$$

- 3 Recevoir la réponse $f(x)$. En cas d'erreur,

$$\begin{aligned} w_i &\leftarrow w_i + \eta(f(x) - h(x))x_i \\ \theta &\leftarrow \theta + \eta f(x)n^2 \end{aligned}$$



Initialisation

- 1 Choisir les poids w_i et le seuil θ aléatoirement

Tours

- 1 Recevoir une instance x
- 2 Envoyer la prédiction

$$h(x) = \text{sign} \left(\sum_{i=1}^n w_i x_i - \theta \right)$$

- 3 Recevoir la réponse $f(x)$. En cas d'erreur,

$$\begin{aligned} w_i &\leftarrow w_i + \eta(f(x) - h(x))x_i \\ \theta &\leftarrow \theta + \eta f(x)n^2 \end{aligned}$$

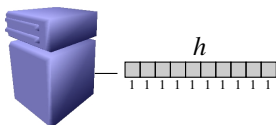
Théorème

Soit δ la marge de la fonction cible. Le nombre maximum d'erreurs que peut faire le perceptron est :

$$O\left(\frac{2n}{\delta}\right)^2$$

Outline

- 1 Cadre Formel
 - Instances
 - Fonctions Linéaires à Seuil
 - Normalisation
 - Séparabilité
- 2 Le perceptron
 - L'algorithme
 - Convergence
- 3 **Winnow**
 - L'algorithme
 - Convergence



Initialisation

- 1 Choisir les poids w_i et le seuil θ à 1

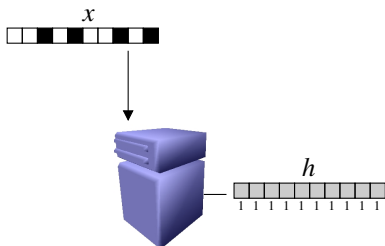
Tours

- 2 Recevoir une instance x
- 3 Envoyer la prédiction

$$h(x) = \text{sign} \left(\sum_{i=1}^n w_i x_i - \theta \right)$$

- 4 Recevoir la réponse $f(x)$. En cas d'erreur,

$$w_i \leftarrow w_i \alpha^{f(x) - h(x)} \alpha_i$$



Initialisation

- 1 Choisir les poids w_i et le seuil θ à 1

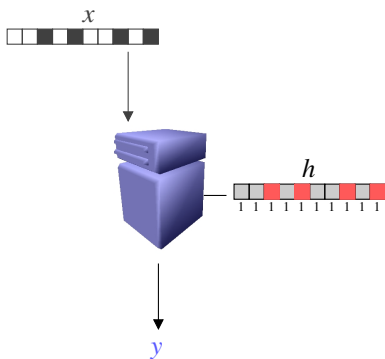
Tours

- 1 Recevoir une instance x
- 2 Envoyer la prédiction

$$h(x) = \text{sign} \left(\sum_{i=1}^n w_i x_i - \theta \right)$$

- 3 Recevoir la réponse $f(x)$. En cas d'erreur,

$$w_i \leftarrow w_i \alpha^{(f(x) - h(x)) x_i}$$



Initialisation

- 1 Choisir les poids w_i et le seuil θ à 1

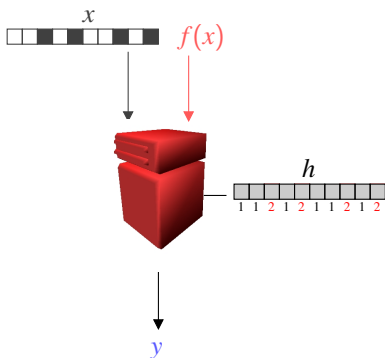
Tours

- 1 Recevoir une instance x
- 2 Envoyer la prédiction

$$h(x) = \text{sign} \left(\sum_{i=1}^n w_i x_i - \theta \right)$$

- 3 Recevoir la réponse $f(x)$. En cas d'erreur,

$$w_i \leftarrow w_i \alpha^{(f(x) - h(x)) x_i}$$



Initialisation

- 1 Choisir les poids w_i et le seuil θ à 1

Tours

- 1 Recevoir une instance x
- 2 Envoyer la prédiction

$$h(x) = \text{sign} \left(\sum_{i=1}^n w_i x_i - \theta \right)$$

- 3 Recevoir la réponse $f(x)$. En cas d'**erreur**,

$$w_i \leftarrow w_i \alpha^{(f(x) - h(x))x_i}$$

Théorème

Soit δ la marge de la fonction cible. Le nombre maximum d'erreurs que peut faire Winnow est :

$$O\left(\frac{\log_2(n)}{\delta^2}\right)$$