Convergence de l'algorithme du perceptron

On considère un perceptron correspondant à une fonction $f:\mathbb{R}^p \to \{-1,1\}$ déterminée par

$$f(x_1,\dots,x_p) = \begin{cases} 1 & \text{si } \omega_1 x_1 + \cdots + \omega_p x_p + \theta \geq 0 \\ -1 & \text{sinon} \end{cases}$$

Rq : La fonction est à valeurs dans $\{-1,1\}$ au lieu de $\{0,1\}$ pour faciliter l'exposé, mais cela ne change rien au fond de l'étude…

On se donne un ensemble d'apprentissage

$$A = \{(x^1, y^1), \dots, (x^N, y^N)\}$$

avec pour chaque $i \in \{1,\dots,N\}$, $x^i = (x_1^i, \dots, x_p^i) \in \mathbb{R}^p$ et $y^i \in \{1,-1\}$ la réponse attendue par le perceptron.

Déf : On dit que la base d'apprentissage A est linéairement séparable s'il existe des coefficients $(\omega_1,\dots,\omega_p) \in \mathbb{R}^p \setminus \{(0,\dots,0)\}$ et $\theta \in \mathbb{R}$ tel que

$$\forall 1 \leq i \leq N, y^i = \operatorname{sgn}(\omega_1 x_1^i + \cdots + \omega_p x_p^i + \theta)$$

en posant

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x = 0 \\ -1 & \text{si } x < 0 \end{cases}$$

Pour alléger les notations, on va poser $\theta = \omega_0$ et convenir $x_0^i = 1$ pour chaque $i \in \{1,\dots,N\}$ de sorte que les expressions $\omega_1 x_1 + \cdots + \omega_p x_p + \theta$ s'écrivent plus légèrement

$$\omega_0 x_0 + \omega_1 x_1 + \cdots + \omega_p x_p$$

La définition du perceptron se ramène alors la détermination d'un vecteur synaptique

$$\omega = (\omega_0, \omega_1, \dots, \omega_p) \in \mathbb{R}^{p+1}$$

L'algorithme d'apprentissage du perceptron est alors le suivant :
On choisit arbitrairement un vecteur synaptique $\omega^0 \in \mathbb{R}^{n+1}$.
Tant que les réponses produites ne sont pas toutes correctes :
 on choisit un exemple (x^i, y^i) dans la base d'apprentissage.
 Si la réponse produite par le perceptron sur cet exemple est incorrecte Alors
  on modifie chaque poids synaptiques par la relation :

$$\omega_j^{t+1} := \omega_j^t + y^i x_j^i$$

 Fin Si
Fin Tant que

Théorème :

Si la base d'apprentissage est linéairement séparable, cet algorithme s'arrête et détermine donc un perceptron solution.

dém. :

Puisque la base est linéairement séparable, il existe un vecteur $\omega^* = (\omega_0^*, \omega_1^*, \dots, \omega_p^*)$ non nul tel que

$$\forall 1 \le i \le N, y_i = \operatorname{sgn}(\omega_0^* + \omega_1^* x_1^i + \cdots + \omega_p^* x_p^i)$$

Pour alléger les notations, introduisons $(.|.)$ le produit scalaire canonique sur $\mathbb{R}^{p+1}$. On a donc

$$\forall 1 \le i \le N, y_i = \operatorname{sgn}\left(\omega^* \mid x^i\right)$$

avec rappelons-le $x^i = (1, x_1^i, \dots, x_p^i)$.

Posons

$$\delta = \min_{1 \le i \le n} \left|\left(\omega^* \mid x^i\right)\right| > 0 \text{ et } M = \max_{1 \le i \le n} \left\|x^i\right\| \in \mathbb{R}^+$$

Si par l'absurde, l'algorithme précédent ne s'arrête pas, il construit une suite infinie de vecteurs $(\omega^t)_{t \in \mathbb{N}}$ avec la propriété

$$\omega^{t+1} = \omega^t + y^i x^i$$

lorsque la réponse du percepton défini par le vecteur synaptique ω^t n'est pas bonne sur l'exemple (x^i, y^i).

D'une part

$$\left(\omega^{t+1} \mid \omega^*\right) = \left(\omega^t \mid \omega^*\right) + y^i \left(x^i \mid \omega^*\right) = \left(\omega^t \mid \omega^*\right) + \operatorname{sgn}\left(\omega^* \mid x^i\right)\left(\omega^* \mid x^i\right)$$

donc

$$\left(\omega^{t+1} \mid \omega^*\right) = \left(\omega^t \mid \omega^*\right) + \left|\left(\omega^* \mid x^i\right)\right| \ge \left(\omega^t \mid \omega^*\right) + \delta$$

puis par récurrence

$$\left(\omega^t \mid \omega^*\right) \ge \left(\omega^0 \mid \omega^*\right) + t\delta \qquad (1)$$

D'autre part

$$\left\|\omega^{t+1}\right\|^2 = \left\|\omega^t\right\|^2 + 2y^i \left(\omega^t \mid x^i\right) + \left\|x^i\right\|^2$$

Or $y^i(\omega^t \mid x^i) \le 0$ car la réponse fournie par le perceptron donné par le vecteur synaptique ω^t est supposée incorrecte sur l'exemple (x^i, y^i) et donc

$$\left\|\omega^{t+1}\right\|^2 \le \left\|\omega^t\right\|^2 + \left\|x^i\right\|^2 \le \left\|\omega^t\right\|^2 + M^2$$

puis par récurrence

$$\left\|\omega^t\right\|^2 \le \left\|\omega^0\right\|^2 + tM^2 \qquad (2)$$

Par l'inégalité de Cauchy-Schwarz

$$\left|(\omega^t \mid \omega^*)\right| \le \left\|\omega^t\right\| \left\|\omega^*\right\|$$

donc (1) et (2) donnent

$$\left(\omega^0 \mid \omega^*\right) + t\delta \le \sqrt{\left\|\omega^0\right\|^2 + tM^2}\, \left\|\omega^*\right\|$$

ce qui est incompatible avec un passage à la limite à l'infini : par exemple en divisant par t avant de faire tendre t vers $+\infty$, on obtient $\delta \le 0$ qui est contraire aux hypothèses de travail.