Introduction à l'Apprentissage Automatique – séance 1

1 Perceptron

1. Exécuter l'algorithme de perceptron sur l'ensemble d'apprentissage suivant :

$$S = \left\{ \left(x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, y_1 = +1 \right), \left(x_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, y_2 = -1 \right), \left(x_3 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}, y_3 = -1 \right) \right\},$$

   après avoir placé les points, et fait les transformations nécessaires pour que l'algorithme du perceptron puisse être mis en œuvre (justifier les choix effectués). Représenter la surface de décision obtenue.
2. Soit $S = \{(x_i, y_i)\}_{i=1}^n$ un ensemble d'apprentissage tel que :
   – $\exists w^*, \gamma > 0 : \|w^*\|_2 = 1$ et $y_i\langle w^*, x_i\rangle \geq \gamma, i = 1, \ldots, n$;
   – $\exists R > 0 : \|x_i\| \leq R, i = 1, \ldots, n$.
   Montrer que l'algorithme d'apprentissage du perceptron fait au plus R^2/γ^2 mises à jour.

2 Perceptron modifié

Dans cet exercice, les exemples appartiennent à l'ensemble $\mathbb{R}^n$, où n est un entier fixé et on suppose qu'ils sont tous de **norme 1**. Soit l'algorithme d'apprentissage suivant, prenant un paramètre $\sigma > 0$ et un ensemble d'apprentissage $\mathcal{S} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_\ell, y_\ell)\}$ (avec $y_i \in \{-1, +1\}$) en entrée :

`Perceptron modifié`**:**

1. Choisir aléatoirement (de façon uniforme) un vecteur unitaire $\mathbf{w}$ de $\mathbb{R}^n$
2. **Si** tous les points mal classés (i.e. ceux qui vérifient $y_i\mathbf{w}\cdot\mathbf{x}_i < 0$) de $\mathcal{S}$ vérifient $|w\cdot\mathbf{x}| \leq \sigma\|w\|$ **alors**
   – arrêter
3. **sinon**
   – choisir le point $(\mathbf{x}, y) \in \mathcal{S}$ *mal classé* qui maximise la quantité $|\mathbf{w}\cdot\mathbf{x}|/\|\mathbf{w}\|$
   – mettre $\mathbf{w}$ à jour selon :
   $$\mathbf{w} \leftarrow \mathbf{w} - (\mathbf{w}\cdot\mathbf{x})\mathbf{x}$$
4. **Si** moins de $(1/\sigma^2)\ln n$ mises à jour de $\mathbf{w}$ ont été faites, **alors**
   – retourner au point 2
5. **sinon**
   – retourner au point 1

Nous allons étudier le temps de convergence de cet algorithme. Pour cela, on suppose qu'il existe un hyperplan cible défini par le vecteur unitaire $\mathbf{w}^*$ permettant la séparation de $\mathcal{S}$.

1. Supposons que le vecteur $\mathbf{w}$ unitaire choisi aléatoirement au point 1 de l'algorithme vérifie $\mathbf{w}^*\cdot\mathbf{w} \geq 1/\sqrt{n}$:
   (a) Montrer qu'à chaque mise à jour $\mathbf{w}'$ de $\mathbf{w}$ (cf. point 3) on a :
   $$\mathbf{w}'\cdot\mathbf{w}^* \geq \mathbf{w}\cdot\mathbf{w}^*$$
   (b) Montrer que, de plus,
   $$\|\mathbf{w}'\|^2 \leq \|\mathbf{w}\|^2(1-\sigma^2)$$
   (c) En déduire qu'au bout de t mises à jour, le vecteur $\mathbf{w}$ courant vérifie :
   $$\|\mathbf{w}\| \leq (1-\sigma^2)^{t/2}$$

(d) En déduire que le nombre d'itérations t vérifie nécessairement :

$$t \leq \frac{1}{\sigma^2} \ln n$$

Il vient d'être montré que si jamais le $\mathbf{w}$ inital vérifie bien $\mathbf{w}^* \cdot \mathbf{w} \geq 1/\sqrt{n}$ alors l'algorithme produit un vecteur qui vérifie la condition 2 en un nombre de mises à jour inférieur à $\frac{1}{\sigma^2} \ln n$.

Note : tous les calculs mis en jeu dans cette question son très proches de ceux vus en cours pour la convergence du perceptron usuel.

2. La probabilité pour qu'un vecteur $\mathbf{w}$ choisi uniformément dans la boule unité vérifie $\mathbf{w}^* \cdot \mathbf{w} \geq 1/\sqrt{n}$ est supérieur à $\frac{1}{8}$.
   - Quelle est la probabilité pour que, en faisant deux tirages aléatoires de vecteurs unitaires, aucun d'eux ne vérifie $\mathbf{w}^* \cdot \mathbf{w} \geq 1/\sqrt{n}$?
   - Quelle est la probabilité pour que, en faisant m tirages de vecteurs unitaires, aucun d'eux ne vérifie cette même condition ?
   - Le paramètre δ étant donné, déduire de la question précédente le nombre de tirages de vecteurs unitaires à faire pour être sûr avec une probabilité $1 - \delta$, que l'un d'eux au moins vérifie la condition voulue.
3. Déduire des questions précédentes, en fonction de δ, σ et n que le temps d'exécution de l'algorithme du perceptron modifié pour produire avec une probabilité $1 - \delta$ un vecteur satisfaisant la condition d'arrêt 2 est de l'ordre de $\frac{1}{\sigma^2} \ln n \ln \frac{1}{\delta}$.