Perceptron

Modèles supervisé pour Data Science

Exercice 1.

On considère un modèle de type perceptron linéaire à seuil sans couche cachée et une cellule de sortie, opérant sur des données de dimension 2. Ce perceptron permet de réaliser des tâches de discrimination / classification à deux classes, notées C_1 et C_2. On note $\mathbf{w}$ le vecteur de poids du perceptron et $\mathbf{x}$ un vecteur représentant une forme d'entrée à classifier entre deux classes.

1.1 Exprimer la sortie calculée par le modèle en fonction de ses entrées.En général, on introduit un terme constant dans la fonction calculée par le réseau. Pour donner lieu à une modélisation simple, on formalise ce terme constant par une entrée constante égale à 1, artificiellement ajoutée aux vecteurs de forme $\mathbf{x}$, ce qui rajoute un poids au vecteur $\mathbf{w}$.

1.2 Ré-exprimer la sortie calculée par le réseau en fonction de ses entrées (On notera w_0 le poids associé à cette entrée particulière).

1.3 Quel est l'intérêt de ce terme constant, appelé biais? Dans la suite une forme $\mathbf{x}$ inclura d'office cette composante additionnelle à 1.On pose que les deux classes C_1 et C_2 sont codées par des valeurs $+1$ et -1. Le perceptron doit donc produire, pour une forme $\mathbf{x}$ en entrée, une valeur de sortie $+1$ si $\mathbf{x}$ est de la classe C_1 et -1 si $\mathbf{x}$ est de la classe C_2.

En apprentissage supervisé, on dispose d'un ensemble de formes d'apprentissages (i.e. de vecteurs $\mathbf{x}$) étiquetés, c.a.d. d'exemples dont on connaît la classe (C_1 ou C_2). Le but de l'apprentissage est de déterminer les paramètres du perceptron ($\mathbf{w}$) qui permettent de respecter (au mieux) l'information de classe contenue dans les données d'apprentissage.

1.4 À quelle condition (formule faisant intervenir la forme d'entrée, $\mathbf{x}$, y, sa classe, et le vecteur de poids $\mathbf{x}$) un exemple est-il bien classé?

1.5 Le réseau implémente une frontière de décision entre les classes. Quelle est sa forme? Quelle est son équation?On considère le problème du ET logique à deux arguments. Les entrées sont des couples de valeurs 1 ou -1 (Vrai ou Faux) et les sorties sont des valeurs 1 et -1 (Vrai ou Faux).

1.6 Donner par extension l'ensemble des formes et leurs classes? Trouver les paramètres du perceptron qui permettent de réaliser cette tâche. Est-il unique?

1.7 Dessiner, dans le plan (en omettant le composante à 1), les formes des différentes classes ainsi que la frontière de décision implémentée par le réseau trouvé.

1.8 Répondre aux trois premières questions du 4. pour la fonction OU exclusif. Idem. Pas forcément à faire ou alors en accéléré. Ne pas tout dessiner par exemple.

EXERCICE 2.

Nous supposons qu'il existe un vecteur $\mathbf{w}^*$ qui discrimine parfaitement tous les exemples d'apprentissage $S = \{(\mathbf{x}_j, y_j), j \in \{1, \ldots, m\}\}$, et nous définissons

$$\rho = \min_{j \in \{1,\ldots,m\}} \left(y_j \left\langle \frac{\boldsymbol{w}^*}{||\boldsymbol{w}^*||}, \mathbf{x}_j \right\rangle \right).$$

2.1 À quoi correspond ρ? Expliquer pourquoi elle serait une valeur strictement positive.

Nous supposons de plus que tous les exemples d'apprentissage sont contenus dans un hypersphère de rayon R (i.e. $\forall j, ||\mathbf{x}_j|| \leq R$). Sans restreindre la généralité, nous supposons de plus que le vecteur poids initial vaut $\boldsymbol{w}^{(0)} = 0$, et nous initialisons le pas d'apprentissage ϵ à 1.

2.2 Montrer qu'après l mises à jour, la norme du vecteur poids vérifie

$$||\boldsymbol{w}^{(l)}||^2 \leq l \times R^2 \tag{1}$$

2.3 Avec les mêmes conditions que précédemment, montrer qu'après l mises du vecteur poids nous avons

$$\left\langle \frac{\boldsymbol{w}^*}{||\boldsymbol{w}^*||}, \boldsymbol{w}^{(l)} \right\rangle \geq l \times \rho \tag{2}$$

Déduire des questions précédentes que le nombre de mises à jour, l, de l'algorithme de Peceptron est borné par

$$l \leq \left\lfloor \left(\frac{R}{\rho}\right)^2 \right\rfloor$$

Où $\lfloor x \rfloor$ est la partie entière du réel x.