TD 5 et 6

Exercice 1 – Support Vector Machine

On considère ici un problème de classification binaire vers $Y = \{-1, +1\}$ de données dans un espace de description $X \in \mathbb{R}^d$. On note $\{(\mathbf{x}^i, y^i) \in (X, Y)\}, i \in \{1, \dots, n\}$ l'ensemble d'apprentissage considéré. La fonction de décision du classifieur considéré est donnée par : $f_{\mathbf{w},b}(\mathbf{x}) = sign(\mathbf{w}^T\mathbf{x} + b)$.
On considère dans un premier temps un ensemble de données linéairement séparable. Cet ensemble de données et la frontière de décision sont représentés (en 2D) sur la figure 1.

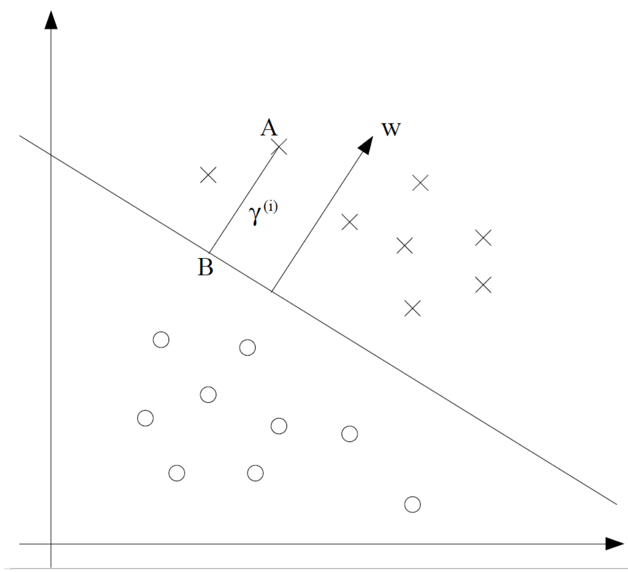

FIGURE 1 – Ensemble de données linéairement séparables

Q 1.1 Marge
Sur cette figure, l'échantillon $\mathbf{x}^i$ et de label y^i est représenté par le point A. On s'intéresse à sa distance signée γ^i à la frontière de decision (dont le point le plus proche est représenté en B sur la figure).

Q 1.1.1 Sachant que $\mathbf{w}/||\mathbf{w}||$ est un vecteur unitaire othogonal à la frontière de décision, donner l'expression de γ^i en fonction de $\mathbf{x}^i$, y^i, $\mathbf{w}$ et b.

Q 1.1.2 Montrer que la distance et la solution ne change pas en multipliant la solution par un scalaire, i.e. pour $(\alpha\mathbf{w}, \alpha b)$. Que cela implique-t-il si l'on souhaite éloigner au maximum (au sens géométrique) les points de la frontière de décision ?

Q 1.2 Formulation du SVM
On considère alors le problème d'optimisation sous contraintes suivant :

$$\min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2$$

$$s.t. \qquad y^i(\mathbf{w}^T\mathbf{x}^i + b) \geq 1, \forall i \in \{1, \dots, n\}$$

Q 1.2.1 Pourquoi choisit-on la contrainte ≥ 1 plutôt que ≥ 0 ? Pourquoi 1 ?

Q 1.2.2 Poser le Lagrangien à considérer pour optimiser ce problème sous contraintes

Q 1.2.3 Donner la solution analytique de la minimisation de ce Lagrangien selon $\mathbf{w}$ et b

Q 1.2.4 En déduire une nouvelle formulation “duale” de notre problème d’optimisation sous contraintes

Q 1.2.5 Que cette nouvelle formulation permet-elle ?

Q 1.2.6 Quel est le problème du problème d’optimisation que l’on a considéré ? Proposer une nouvelle formulation qui corrige ce problème

Q 1.2.7 Proposer la formulation duale de ce nouveau problème

Q 1.2.8 Donner la fonction de classification obtenue après optimisation de cette formulation duale du SVM

Q 1.2.9 D’après les conditions de Karush-Kuhn-Tucker (KKT) concernant les propriétés de la solution optimale d’un Lagrangien, on a : $a_i(1-\xi_i-y^i(\mathbf{w}^T\mathbf{x}^i+b))=0,\ \forall i\in\{1..N\}$ et $\beta_i\xi_i=0,\ \forall i\in\{1..N\}$. Qu’en déduire pour les paramètres a_i obtenus à l’optimum ?

Q 1.2.10 Qu’en déduire pour l’estimation du biais b ?

Exercice 2 – Noyaux

Q 2.1 Montrez que si K et K' sont deux noyaux (i.e. il existe ϕ et ϕ' telles que $K(x,y)=<\phi(x),\phi(y)>$, $K'(x,y)=<\phi'(x),\phi'(y)>$) :

Q 2.1.1 cK est un noyau pour $c\in\mathbb{R}^+$

Q 2.1.2 $K+K'$ est un noyau ;

Q 2.1.3 KK' est un noyau ;

Q 2.1.4 $(1+<x,x'>)^d$ est un noyau.

Exercice 3 – RKHS

Soit $x_1,\cdots,x_n\in X$, une fonction $k:X\times X\to\mathbb{R}$, la matrice de Gram de K est la matrice $\mathcal{K}:=k_{i,j}=k(x_i,x_j)$. Une matrice est dite définie semi-positive si $\forall c_i\in\mathbb{R},\sum_{i,j}c_ic_jk_{i,j}\geq 0$. Dans ce cas, la fonction est dite également définie positive.

Q 3.1 Exprimez $\sum_{i,j}c_ic_jk_{i,j}$ par un produit scalaire. Montrez qu’un noyau est défini positif.

Q 3.2 Le but de cette question est de montrer la contraposée, qu’une fonction symétrique semi définie positive $k:X\times X\to\mathbb{R}$ est un noyau. Pour cela, il nous faut trouver un espace hilbertien $\mathcal{H}$, un produit scalaire $Q:\mathcal{H}\times\mathcal{H}\to\mathbb{R}$ et une projection $\phi:X\to\mathcal{H}$ telle que $k(x,y)=Q(\phi(x),\phi(y))\ \ \forall x,y\in X$. On va considérer $\mathcal{H}$ l’espace vectoriel engendré par les fonctions de la forme $y\to k(y,x)$ pour tout $x\in X$. Un élément de $\mathcal{H}$ est donc une fonction de $X\to\mathbb{R}$.
Soit $\Phi:X\to\mathcal{H}\ :=k(.,x)$ un mapping de X aux fonctions de $\mathcal{H}$, $\Phi(x)(x')=k(x',x)$. Soient $\alpha_i\in\mathbb{R}$, $\beta_i\in\mathbb{R}$, $x_i\in X$, $x'_i\in X$ pour $i\in\{1..n\}$. On définit :

$$f(.)=\sum_{i=1}^n\alpha_i\Phi(x_i)(.),\quad g(.)=\sum_{i=1}^n\beta_i\Phi(x'_i)(.),\quad Q(f,g)=\sum_{i=1}^n\sum_{j=1}^n\alpha_i\beta_jk(x_i,x'_j)$$

f et g sont bien dans $\mathcal{H}$, vu que ce sont des combinaisons linéaires d’éléments de $\mathcal{H}$.

Q 3.2.1 Montrez que Q peut s’exprimer uniquement à l’aide des β_j et $f(x'_j)$, ou des α_i et $g(x_i)$

Q 3.2.2 Montrez que $Q(f,g)$ est un produit scalaire de f et g (on pourra alors remplacer $Q(f,g)$ par $\langle f,g\rangle$). Pour cela, il s’agit de démontrer que :

- Q est symétrique

- Q est bilinéaire
- $Q(f, f) \geq 0$ (on montrera dans la dernière question que $Q(f, f) = 0 \iff f = 0$).

Q 3.2.3 Que vaut $Q(k(., x), f)$? $Q(k(., x), k(., x'))$? Justifiez le nom de k : *reproducing kernel.*

Q 3.2.4 En admettant que $Q(f, g)^2 \leq Q(f, f)^2 Q(g, g)^2$, montrez que $|f(x)|^2 \leq k(x, x).Q(f, f)$. Concluez.

Exercice 4 – Noyaux sur les chaînes de caractères

Soit S une séquence de mots sur un alphabet $\mathcal{A}$ fini. Montrez que :

1. $K(x, x')$ = nombre de sous-chaînes de longueur 5 que x et x' ont en commun est un noyau ;
2. $K(x, x') = 1$ si x et x' ont au moins une sous-chaîne de longueur 5 en commun, 0 sinon, n'est pas un noyau (indice : considérez 3 chaînes x,x' et x'').