TD 1

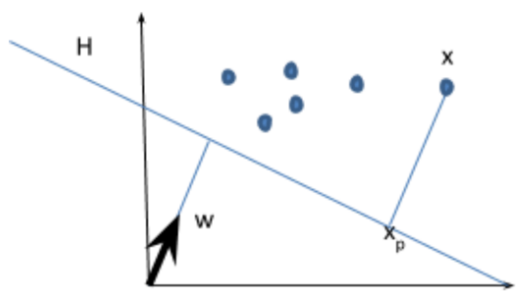
Exercice 1 – One Class SVM

On dispose d'un ensemble de données non étiquetées, on s'intéresse à construire un système qui réponde $+1$ sur des données proches de cet ensemble d'apprentissage et -1 ailleurs. Pensez à un problème de classification où l'on ne dispose que d'exemples positifs et pas d'exemple négatifs. On considère un ensemble de données $\{\mathbf{x}^1, \cdots, \mathbf{x}^N\}$, avec $\mathbf{x}^i \in \mathbb{R}^d$, on va définir un classifieur à marge qui répond $+1$ dans un voisinage de ces données et -1 partout ailleurs. Pour cela, on va résoudre le problème suivant (pb 1) :

$$\min_{w,\rho} \frac{1}{2}||\mathbf{w}||^2 - \rho \text{ sous les contraintes } \mathbf{w}.\mathbf{x}^i \geq \rho, \quad \rho \in \mathbb{R}$$

.

Q 1.1 Soit la figure suivante :



Q 1.1.1 Soit H l'hyperplan d'équation $F(\mathbf{x}) = \mathbf{w}.\mathbf{x} - \rho = 0$. Pour un point $\mathbf{x}$, on note x_p sa projection sur l'hyperplan H. Montrer que la distance de $\mathbf{x}$ à H s'écrit $\frac{F(\mathbf{x})}{||\mathbf{w}||}$ (en écrivant $\mathbf{x}$ sous la forme $\mathbf{x} = x_p + r\frac{\mathbf{w}}{||\mathbf{w}||}$ où r est la distance recherchée. En déduire que la distance de l'orginie à H vaut $\frac{|\rho|}{||\mathbf{w}||}$.

Q 1.1.2 On considère la fonction de décision : $G(\mathbf{x}) = sign(F(\mathbf{x}))$. Positionner sur la figure un hyperplan qui permet de résoudre le problème (pb1).

Q 1.1.3 Cet hyperplan est-il unique ? Justifier.

Q 1.2 Ecrire le lagrangien $L(\mathbf{w}, \rho, \alpha)$ associé au problème (pb1).

Q 1.3 Calculer les dérivées $\frac{\partial L}{\partial w_j}$ et $\frac{\partial L}{\partial \rho}$.

Q 1.4 En déduire $\mathbf{w} = \sum_{i=1}^N \alpha_i \mathbf{x}^i$ et $\sum_{i=1}^N \alpha_i = 1$.

Q 1.5 Montrer que la formulation duale du problème (pb1) s'écrit (pb2) :

$$min_\alpha \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j \mathbf{x}^i \mathbf{x}^j \quad \text{ sous les contraintes } \alpha_i \geq 0, \ \sum_{i=1}^N \alpha_i = 1$$

Q 1.6 Montrer qu'à l'optimum seuls les points pour lesquels $\mathbf{w}.\mathbf{x}^i = \rho$ ont un coefficient α_i non nul.

Q 1.7 Montrer que $\rho = \mathbf{w}.\mathbf{x}^i$ pour $\mathbf{x}^i$ vecteur support quelconque.