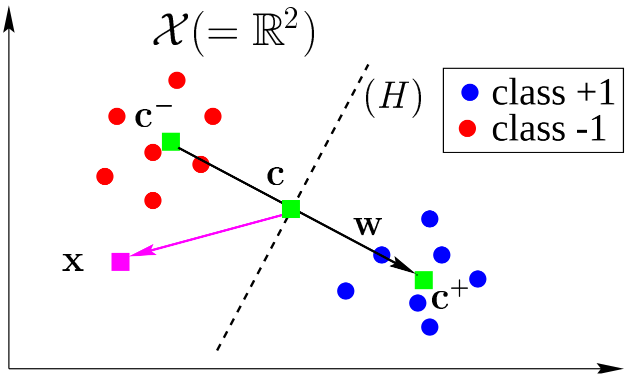
1 Premier classifieur à noyaux

Soit l'un des premiers algorithmes de classification vus en cours qui définit un hyperplan d'équation $\mathbf{w} \cdot \mathbf{x} + b = 0$ en définissant $\mathbf{w}$ comme le vecteur joignant la moyenne des points de classe -1 et celle des points de classe -1 :



où $\mathcal{S} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m))\}$, $\mathbf{x}_i \in \mathcal{X}$ et $y_i \in \mathcal{Y} = \{-1, +1\}$, est l'ensemble d'apprentissage et $\mathcal{S}$ se décompose en m^+ exemples positifs et m^- exemples négatifs.

1. Exprimer $\mathbf{c}^+$ et $\mathbf{c}^-$ en fonction des $\mathbf{x}_i$, m^+ et m^-.
2. Donner une règle simple de classification d'exemples reposant sur $\mathbf{w}$ et b. Donner une expression de $\mathbf{w}$ et de b.
3. Si l'on choisit d'écrire $\mathbf{w}$ sous la forme $\mathbf{w} = \sum_{i=1}^m \alpha_i \mathbf{x}_i$, donner la valeur des coefficients α_i.
4. Supposons qu'on utilise un noyau k à la place du produit scalaire usuelle. Donner une règle de classification utilisant k (règle de classification similaire à celle de la question 2).

2 Perceptron

Soit l'ensemble d'apprentissage

$$\mathcal{S} = \left\{ \left(\mathbf{x}_1 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, +1\right), \left(\mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, -1\right), \left(\mathbf{x}_3 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, +1\right), \left(\mathbf{x}_4 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}, -1\right) \right\}$$

On rappelle l'algorithme d'apprentissage du perceptron à noyau présenté en cours, k étant un noyau, $\mathcal{S} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_\ell, y_\ell)\}$, $\mathbf{x}_i \in \mathbb{R}^n$, $y_i \in \{-1, +1\}$

- $\boldsymbol{\alpha} = \mathbf{0}$
- Répéter jusqu'à convergence ou bien atteinte d'un nombre max d'itérations
  - si i est tel que $y_i \sum_{j=1}^{\ell} y_j \alpha_j k(\mathbf{x}_j, \mathbf{x}_i) \leq 0$
    $\alpha_i \leftarrow \alpha_i + 1$

1. Représenter sur le plan les points d'apprentissage $\mathbf{x}_1, \ldots, \mathbf{x}_4$.
2. Peut-on espérer que l'algorithme du perceptron *linéaire* fournisse une solution au problème d'apprentissage donné ? Justifier.
3. Afin d'aborder ce problème de classification, nous allons utiliser l'algorithme du perceptron à noyau avec $k(u, v) = (\mathbf{u} \cdot \mathbf{v})^2 + 1$.
   (a) proposer une transformation $\phi : \mathbb{R}^2 \to \mathbb{R}^4$ telle que $k(\mathbf{u}, \mathbf{v}) = \langle \phi(\mathbf{u}), \phi(\mathbf{v}) \rangle_{\mathbb{R}^4}$;
   (b) donner la matrice de Gram de k pour l'ensemble $\mathcal{S}$;
   (c) exécuter l'algorithme du perceptron à noyau sur $\mathcal{S}$ (moins d'une dizaine de mises à jour sont nécessaires).
4. En déduire une expression analytique, sous la forme d'une forme quadratique, de la surface de décision ainsi générée (cette expression qui doit être du type $f(x, y) = 0$ vous rappellera les equations de coniques vues en terminale).
5. En fonction de quoi le nombre maximal d'itérations du perceptron à noyau est-il défini ? En déduire le nombre maximal de coefficients α_i du vecteur $\boldsymbol{\alpha}$ qui sont non nuls.

3 Noyaux

3.1 Noyaux basiques

Dans cet exercice, on désigne par $\langle \cdot, \cdot \rangle_d$ l'application

$$\begin{aligned} \langle \cdot, \cdot \rangle_d : \mathbb{R}^d \times \mathbb{R}^d &\to \mathbb{R} \\ (\mathbf{u}, \mathbf{v}) &\mapsto \langle \mathbf{u}, \mathbf{v} \rangle_d = \sum_{i=1}^{d} u_i v_i \end{aligned}$$

1. montrer brièvement que $\langle \cdot, \cdot \rangle_d$ est bilinéaire, symétrique, définie et positive (c'est donc un noyau) ;
2. soit la fonction définie pour un d fixé par $k_d(\mathbf{u}, \mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle_d + 1$; trouver une application ϕ de $\mathbb{R}^d$ dans $\mathbb{R}^{d+1}$ telle que $k(u, v) = \langle \phi(\mathbf{u}), \phi(\mathbf{v}) \rangle_{d+1}$
3. soit l'application :

   $$\begin{aligned} \phi : \mathbb{R}^2 &\to \mathbb{R}^5 \\ \mathbf{u} &\mapsto \phi(\mathbf{u}) = [u_1, u_2, 2u_1u_2, u_1^2, u_2^2]^\top \end{aligned}$$

   simplifier l'écriture du noyau k défini par $k(\mathbf{u}, \mathbf{v}) = \langle \phi(\mathbf{u}), \phi(\mathbf{v}) \rangle_5$;
4. soit $\mathcal{S} = \{\mathbf{x}_1, \ldots, \mathbf{x}_\ell\}$ un ensemble de vecteurs de $\mathbb{R}^d$
   (a) donner la taille et le terme général de la matrice de Gram K associée à $\langle \cdot, \cdot \rangle_d$ et $\mathcal{S}$;
   (b) montrer que les valeurs propres K sont positives ou nulles (indice : utiliser la matrice X de taille $d \times \ell$ définie par $X = [\mathbf{x}_1 \cdots \mathbf{x}_\ell]$)
5. Soit l'application ϕ définie par

   $$\begin{aligned} \phi : \mathbb{R}^2 &\to \mathbb{R}^6 \\ \mathbf{u} &\mapsto \phi(\mathbf{u}) = \left[3u_1, -3u_2, 2u_1u_2, \sqrt{2}u_1^2, \sqrt{2}u_2^2, \sqrt{2}\right]^\top \end{aligned}$$

   simplifier l'écriture du noyau k défini par $k(\mathbf{u}, \mathbf{v}) = \langle \phi(\mathbf{u}), \phi(\mathbf{v}) \rangle_6$ e n faisant en particulier apparaître le produit scalaire $\langle \mathbf{u}, \mathbf{v} \rangle_2$
6. Dans cette question, on suppose que d est un entier naturel fixé. Soit l'application $k_{\cos}$ définie par

   $$\begin{aligned} k_{\cos} : \mathbb{R}^d \times \mathbb{R}^d &\to \mathbb{R} \\ (\mathbf{u}, \mathbf{v}) &\mapsto k_{\cos}(\mathbf{u}, \mathbf{v}) = \cos(\mathbf{u}, \mathbf{v}) \end{aligned}$$

   où $\cos(\mathbf{u}, \mathbf{v})$ désigne le cosinus de l'angle entre les deux vecteurs $\mathbf{u}$ et $\mathbf{v}$. Trouver une application $\phi_{\cos}$ de $\mathbb{R}^d$ dans $\mathbb{R}^d$ telle que

   $$k_{\cos}(\mathbf{u}, \mathbf{v}) = \langle \phi(\mathbf{u}), \phi(\mathbf{v}) \rangle_d$$

3.2 Somme de noyaux de Mercer

1. Etant donné un échantillon $\mathcal{S} = \{\mathbf{x}_1, \ldots, \mathbf{x}_\ell\}$ d'exemples de $\mathcal{X}$, dire ce qu'est la matrice de Gram K d'un noyau de Mercer k défini par rapport à $\mathcal{S}$ (taille de cette matrice, terme général). Donner une propriété caractéristique de cette matrice.
2. Soit k_1 et k_2 deux noyaux de Mercer définis sur $\mathcal{X} \times \mathcal{X}$.
   (a) Etant donné un échantillon $\mathcal{S} = \{\mathbf{x}_1, \ldots, \mathbf{x}_\ell\}$ d'exemples de $\mathcal{X}$, donner la matrice de Gram associée à la fonction $k_1 + k_2$.
   (b) Montrer que puisque k_1 et k_2 sont des noyaux de Mercer, $k_1 + k_2$ est également un noyau de Mercer.