Université de Lorraine
Ecole Nationale Supérieure d'Electricité et de Mécanique

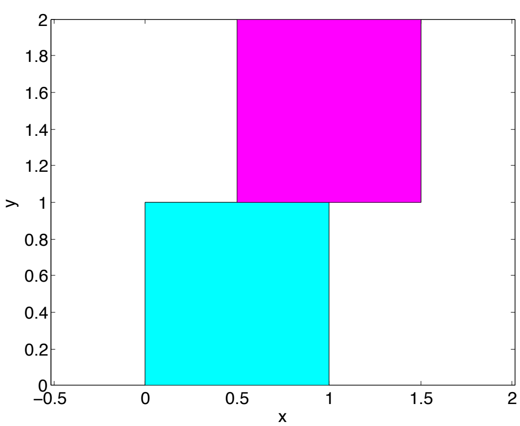
A first look on SVM

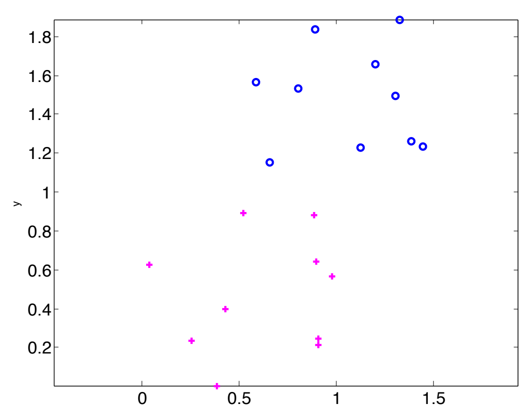
Didier Maquin[1] – Didier.Maquin@univ-lorraine.fr

L'enjeu de cette séance de TD est d'élaborer, en utilisant Matlab®, des programmes implémentant des séparateurs linéaires à vaste marge. On s'intéressera tout d'abord au cas de données linéairement séparables (le problème initial des SVM) en résolvant le problème d'optimisation primal consistant à maximiser la marge. On élaborera ensuite un séparateur linéaire à marge poreuse dans le cas de données non linéairement séparables.

1 Génération des données linéairement séparables

On se propose tout d'abord de générer dans $\mathbb{R}^2$, 2×10 observations séparables linéairement. Pour cela, générer une matrice X_1 de dimension 10×2 dont chaque élément est la réalisation d'une variable aléatoire uniformément distribuée entre 0 et 1 (fonction `rand`). Faire de même pour une seconde matrice X_2 en ajoutant un *offset* égal à 0.5 sur les abscisses et 1 sur les ordonnées. On notera que les observations de X_1 sont donc situées dans le carré défini par les 4 sommets {(0,0), (0,1), (1,1), (1,0)} alors que celles de X_2 sont situées dans le carré défini par {(0.5,1), (0.5,2), (1.5,2), (1.5,1)} comme le montre la figure suivante (à gauche) :





Créer la matrice d'observations $X \in \mathbb{R}^{20\times 2}$ regroupant X_1 et X_2 et créer le vecteur ℓ d'éléments $\ell_k \in \{1, -1\}$ tel que $\ell_k = 1$ pour les observations de X_1 et $\ell_k = -1$ pour les observations de X_2. Visualiser les observations dans $\mathbb{R}^2$ en leur affectant des marqueurs et des couleurs différentes (voir figure ci-dessus à droite).

2 Elaboration du séparateur à vaste marge

On rappelle les résultats suivants.

1. Inspiré de documents pédagogiques de S. Canu (http ://asi.insa-rouen.fr/enseignants/˜scanu/).

Soit $\{(x_k, \ell_k), k = 1, N\}$ un ensemble de vecteurs-observations étiquetés avec $x_k \in \mathbb{R}^p$ et $\ell_k \in \{1, -1\}$. Un séparateur à vaste marge linéaire (SVM) est un discriminateur linéaire de la forme : $D(x) = \text{signe}(w^T x + b)$ où $w \in \mathbb{R}^p$ et $b \in \mathbb{R}$ sont donnés par la résolution du problème suivant :

$$\text{Primal} \quad \begin{cases} \min\limits_{w,b} \dfrac{1}{2}||w||^2 \\ \text{sous } \ell_k(w^T x_k + b) \geq 1, \quad k = 1, \ldots, N \end{cases}$$

Le problème d'optimisation sous contraintes précédent est un “programme quadratique” de la forme générale :

$$\begin{cases} \min\limits_{z} \dfrac{1}{2} z^T H z + f^T z \\ \text{sous } Az \leq e \end{cases}$$

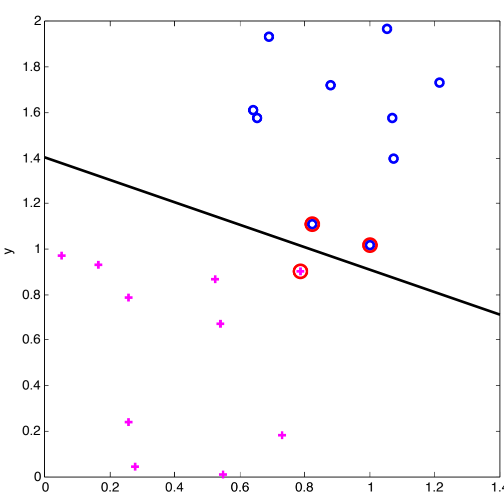
avec

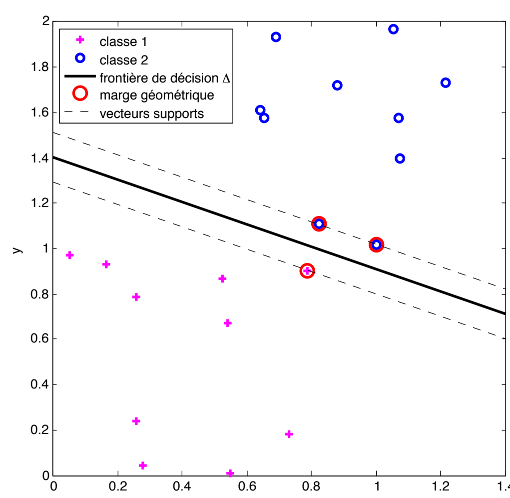
$$z = \begin{pmatrix} w \\ b \end{pmatrix} \in \mathbb{R}^{p+1} \qquad f = (0, \ldots, 0)^T \in \mathbb{R}^{p+1} \qquad H = \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix}$$

$$A = -(\text{diag}(\ell)X \quad \ell) \qquad e = -(1, \ldots, 1)^T \in \mathbb{R}^N \qquad \ell \in \mathbb{R}^N \text{ vecteur des signes}$$

$$X \in \mathbb{R}^{N \times p} \text{ matrice dont la ligne } k \text{ est } x_k^T$$

1. Construire les matrices décrivant le problème d'optimisation à résoudre : H, f, A et e. Lire l'aide de la fonction Matlab® `quadprog` et l'utiliser pour résoudre le problème d'optimisation.
2. Tracer la séparatrice. On rappelle qu'elle est décrite par l'équation $w^T x + b = 0$.
3. La fonction `quadprog` fournit la valeur des paramètres de Lagrange relatifs à chaque contrainte (autant de contraintes que d'observations). On peut donc facilement détecter parmi les observations, les vecteurs supports ; ce sont les observations pour lesquelles les paramètres de Lagrange correspondants sont non-nuls (ou numériquement significativement différent de 0). Détecter ces observations et superposer, sur la représentation graphique un marqueur supplémentaire (figure de gauche ci-dessous).

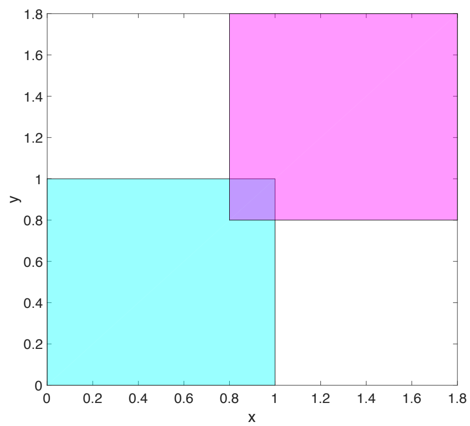




4. Déterminer l'équation des droites parallèles à la séparatrice matérialisant la marge (droites passant par les vecteurs supports et distantes de la séparatrice de $\frac{1}{|w|}$) et les tracer en pointillés.

3 Génération des données non linéairement séparables

Reprendre la méthode de génération précédente en tirant aléatoirement 25 valeurs dans chacun des carrés matérialisés sur la figure suivante.



On notera cependant qu'avec une telle approche, rien ne garantit que les données générées aléatoirement sont bien non-linéairement séparables ! On utilisera donc un jeu de données répondant à cette propriété (ou l'absence de cette propriété).

4 Elaboration du séparateur à marge poreuse

On rappelle que le séparateur peut être obtenu en résolvant le problème d'optimisation suivant :

$$\begin{cases} \min\limits_{w,b,\xi_k} \dfrac{1}{2}\|w\|^2 + C\displaystyle\sum_{k=1}^{N}\xi_k \\ \text{sous} \quad \ell_k(w^T x_k + b) \geq 1 - \xi_k, \quad k = 1,\ldots,N \\ \qquad\quad \xi_k \geq 0, \quad k = 1,\ldots,N \end{cases}$$

Le scalaire C permet de quantifier les poids respectifs de la marge (que l'on cherche à maximiser) et de la somme des variables ressorts (que l'on cherche à minimiser).

1. En s'inspirant du cas précédent, construire les matrices décrivant le problème d'optimisation à résoudre. La fonction `quadprog` peut être utilisée de différentes façons. Le problème peut être ré-écrit sous la forme précédente :

$$\begin{cases} \min\limits_{z} \dfrac{1}{2} z^T H z + f^T z \\ \text{sous } Az \leq e \end{cases}$$

en intégrant dans la matrice A et le vecteur e les contraintes relatives à la positivité des ξ_k. On peut également remarquer que la fonction `quadprog` permet également d'imposer des bornes minimales et maximales sur le vecteur à optimiser. Si l'on n'impose pas de contraintes égalités, l'appel de la fonction peut s'effectuer en utilisant la syntaxe `[x,∼,∼,∼,lambda] = quadprog(H,f,A,b,[],[],lb)`. Le vecteur `lb` doit alors être construit de façon à ne pas imposer de bornes minimales pour w et b (on pourra choisir $-\infty$) et imposer 0 pour chaque ξ_k.

2. Reprendre les questions 2 à 4 du cas des données séparables linéairement. Attention, cette fois, les vecteurs supports correspondent aux observations situées à une distance égale à 1 de la séparatrice. Effectuer différentes simulations en faisant varier le facteur de pondération C.