

LE DATA MINING

La fouille de données

Dr. Nedioui Med Abdelhamid
Université Echahid Hama Lakhder Eloued - Algérie
nediou3904@gmail.com

Chapitre 7

Classification par Machines à Vecteurs de Support SVM

1. Introduction

Le SVM (Machine à Vecteurs de support ou Séparateurs à Vastes Marges) est une méthode de classification par apprentissage supervisé, elle est introduite par Vapnik en 1995. Généralement, les SVM sont considérées comme une approche de classification, mais elles peuvent être utilisées dans les deux types de problèmes de classification et de régression. Il peut facilement gérer plusieurs variables continues et catégorielles. Le SVM construit un hyperplan dans l'espace multidimensionnel pour séparer les différentes classes. L'idée de base de SVM est de trouver un hyperplan marginal maximal qui divise le mieux l'ensemble de données en classes.

2. principe de fonctionnement d'un SVM

2.1 Notions de base : Hyperplan, marge et support vecteur

Pour deux classes d'exemples donnés, le but de SVM est de trouver un classificateur qui va séparer les données et maximiser la distance entre ces deux classes. Avec SVM, ce classificateur est un classificateur linéaire appelé hyperplan.

Il existe en effet une infinité d'hyperplans séparateurs, dont les performances en phase d'apprentissage sont identiques, mais dont les performances en phase de test peuvent être très différentes.

2. Principe de fonctionnement d'un SVM

2.1 Notions de base

2.1.1 Principe

Pour deux classes d'exemples donnés, le but de SVM est de trouver un classificateur qui va séparer les données et maximiser la distance entre ces deux classes. Avec SVM, ce classificateur est un classificateur linéaire appelé hyperplan.

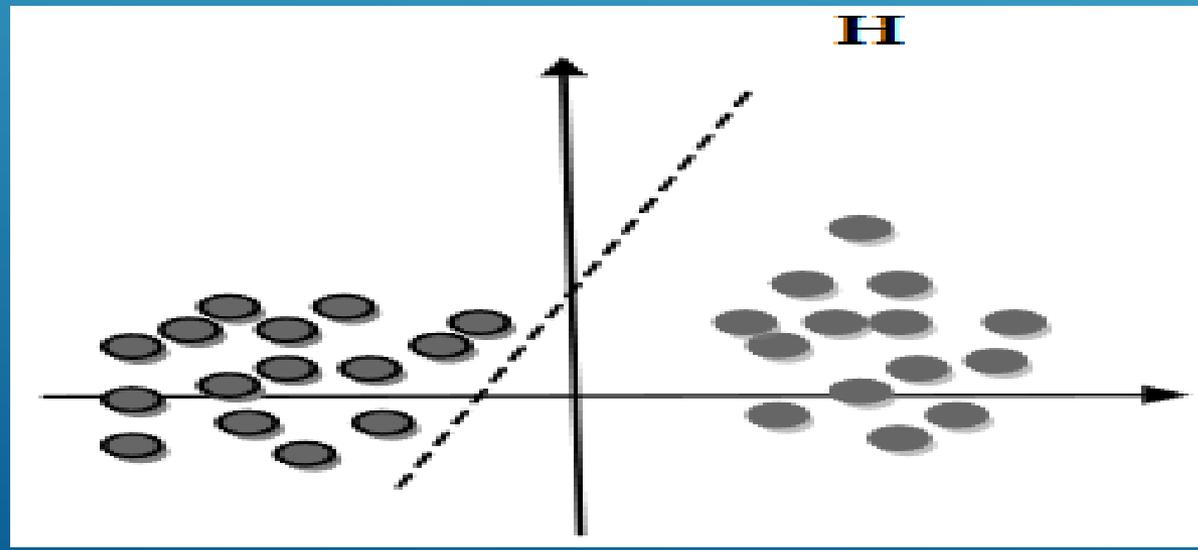
Il existe en effet une infinité d'hyperplans séparateurs, dont les performances en phase d'apprentissage sont identiques, mais dont les performances en phase de test peuvent être très différentes.

2. Principe de fonctionnement d'un SVM

2.1 Notions de base

2.1.2 Hyperplan:

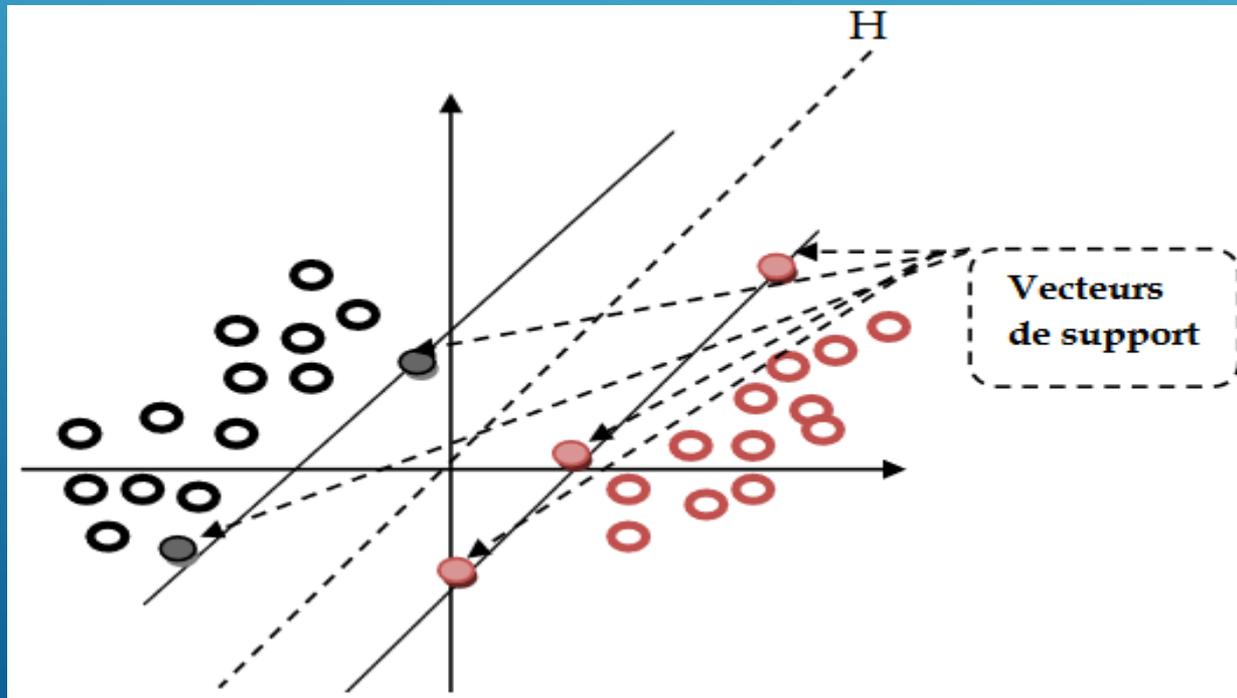
Plaçons-nous dans le cas d'une classification binaire (les exemples à classifier réparties en 2 classes). On appelle **hyperplan séparateur** une ligne qui sépare les deux classes.(il sépare les points d'apprentissage)



2. Principe de fonctionnement d'un SVM

2.1.3 Vecteurs de support

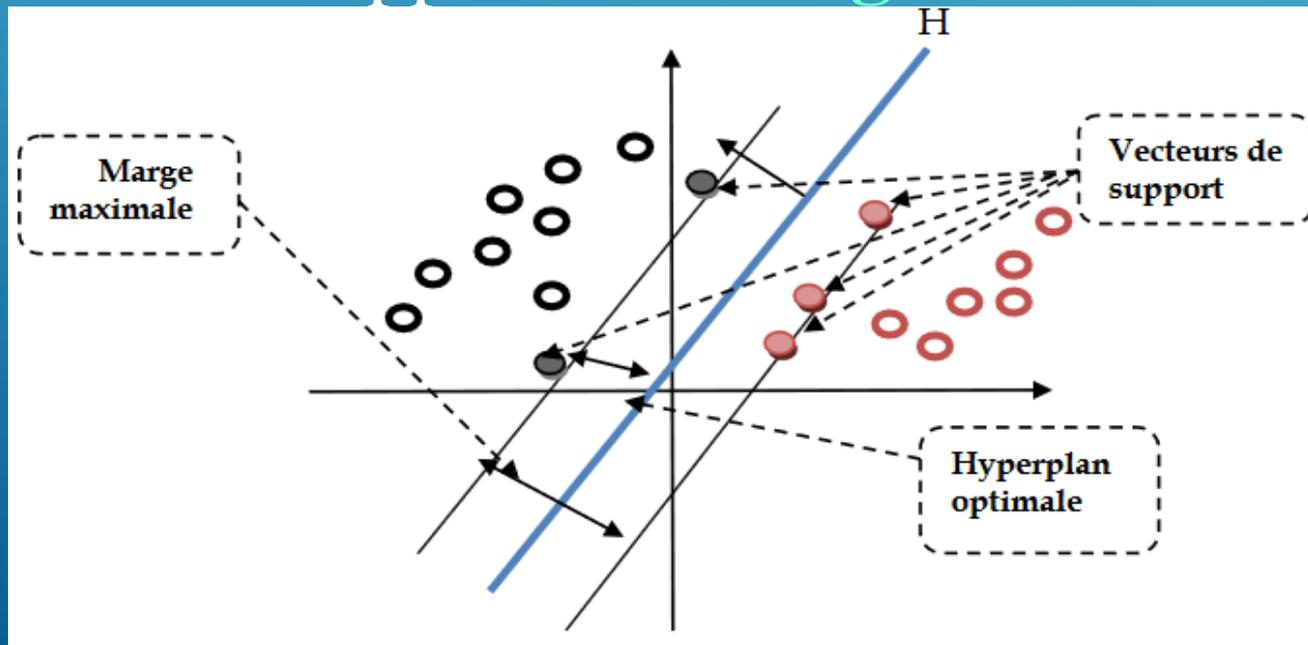
Pour déterminer l'hyperplan séparable des SVM est d'utiliser seulement les points de la frontière entre les deux classes parmi l'ensemble total d'apprentissage, ces points sont appelés **vecteurs de support**



2. Principe de fonctionnement d'un SVM

2.1.4 Marge :

Parmi l'ensemble d'hyperplans capable de séparer parfaitement les deux classes d'exemples, les SVM choisissent celui qui va maximiser la distance minimale entre l'hyperplan et les exemples d'apprentissage, cette distance est appelée la **marge**.



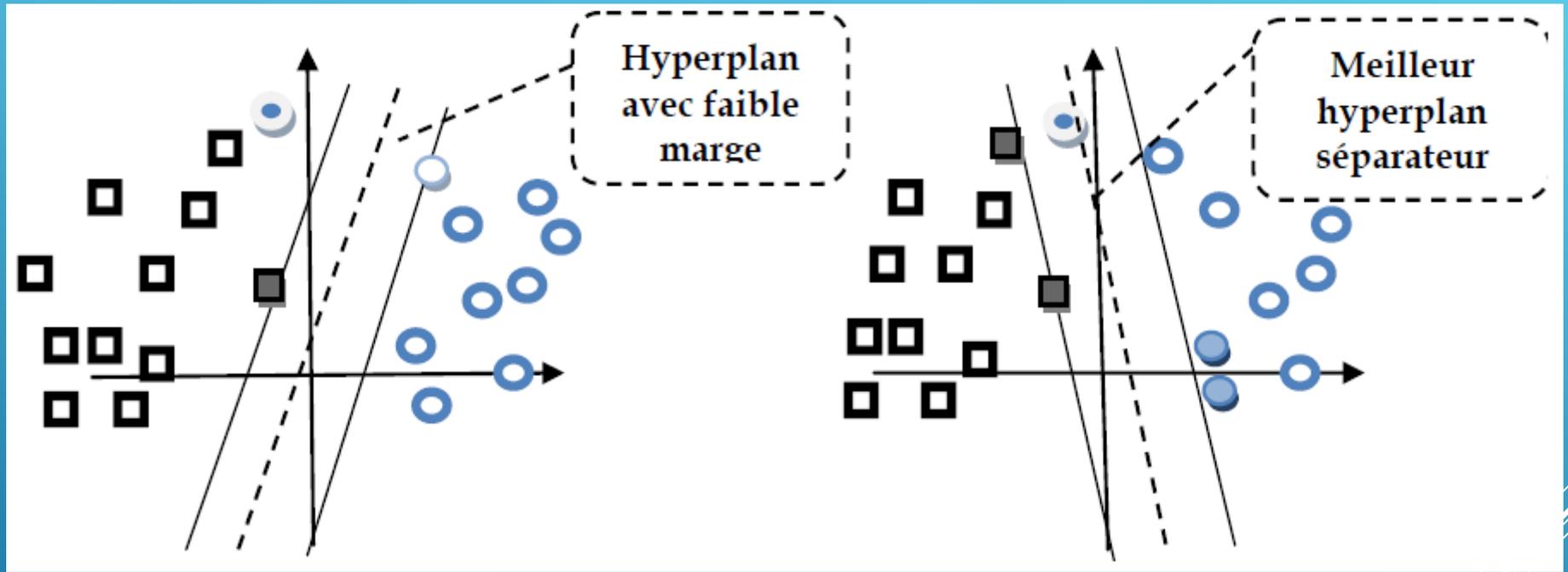
2. Principe de fonctionnement d'un SVM

2.2 Pourquoi maximiser la marge ?

Intuitivement, le fait d'avoir une marge plus large procure plus de sécurité lorsque l'on classe un nouvel exemple. De plus, si l'on trouve le classificateur qui se comporte le mieux vis-à-vis des données d'apprentissage, il est clair qu'il sera aussi celui qui permettra au mieux de classer les nouveaux exemples. Dans le schéma qui suit, la partie droite nous montre qu'avec un hyperplan optimal, un nouvel exemple reste bien classé alors qu'il tombe dans la marge. On constate sur la partie gauche qu'avec une plus petite marge, l'exemple se voit mal classé.

2. Principe de fonctionnement d'un SVM

2.2 Pourquoi maximiser la marge ?



En général, la classification d'un nouvel exemple inconnu est donnée par sa position par rapport à l'hyperplan optimal.

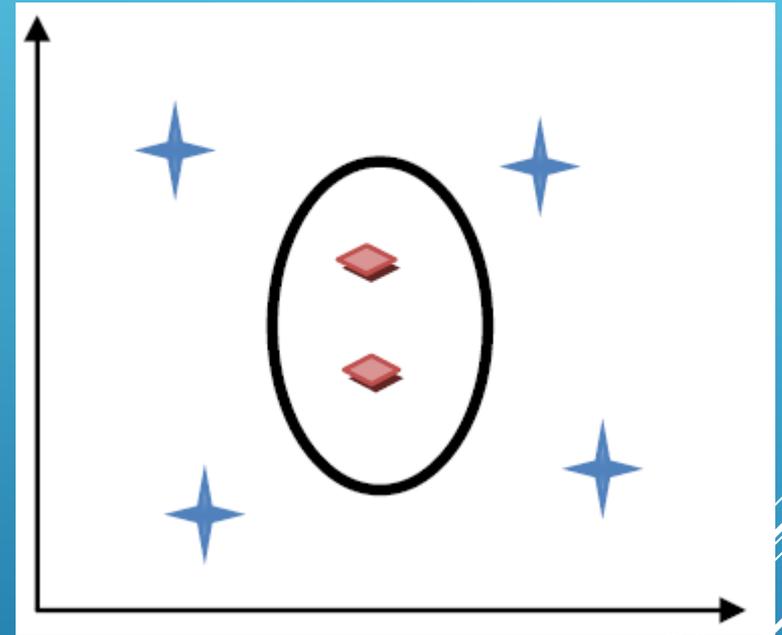
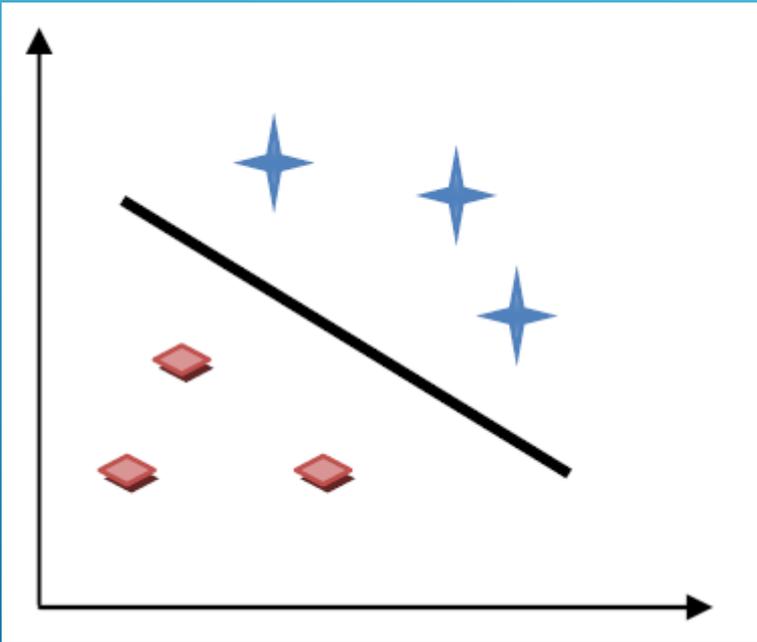
2. Principe de fonctionnement d'un SVM

2.3 Linearite et non-linearite

Parmi les modèles des SVM, on constate les cas linéairement séparable et les cas non linéairement séparable. Les premiers sont les plus simple de SVM car ils permettent de trouver facilement le classificateur linéaire. Dans la plupart des problèmes réels il n'y a pas de séparation linéaire possible entre les données, le classificateur de marge maximale ne peut pas être utilisé car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables.

2. Principe de fonctionnement d'un SVM

2.3 Linearite et non-linearite



2. Principe de fonctionnement d'un SVM

2.4 Cas non linéaire

Pour surmonter les inconvénients des cas non linéairement séparable, l'idée des SVM est de changer l'espace des données. La transformation non linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace. On va donc avoir un changement de dimension. La nouvelle dimension est appelé **espace de re-description**.

En effet, plus la dimension de l'espace de re-description est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples est élevée.

2. Principe de fonctionnement d'un SVM

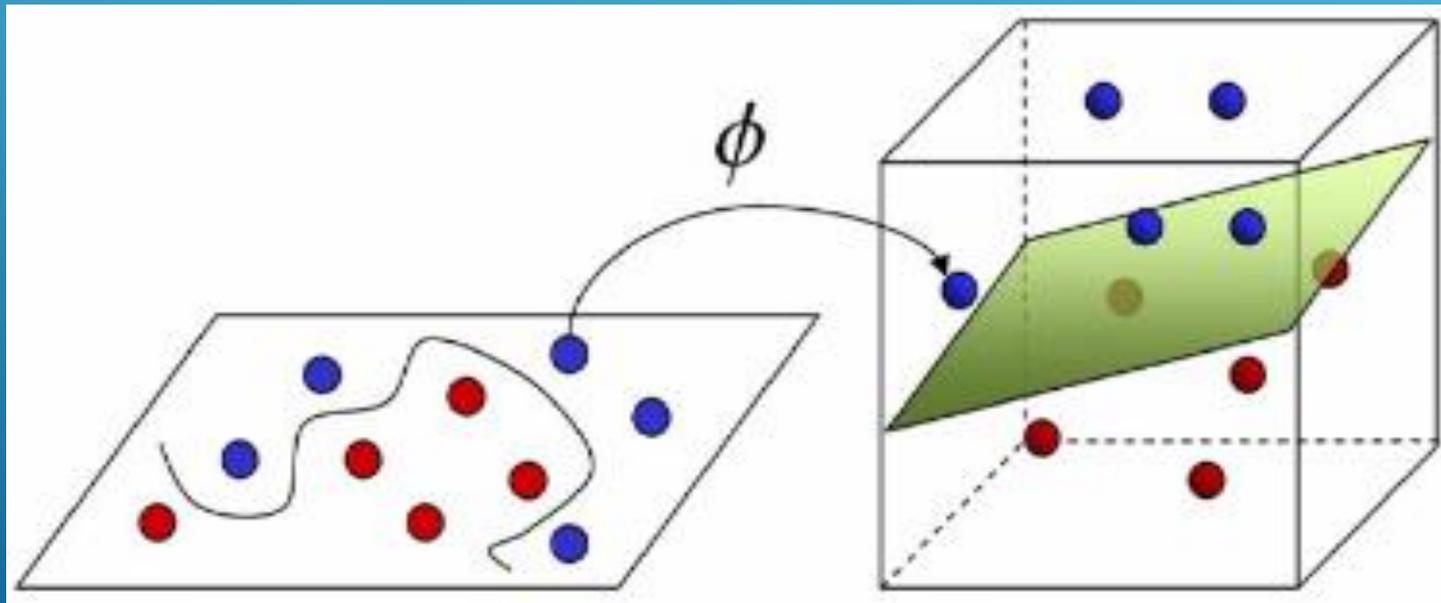
2.4 Cas non linéaire

On a donc une transformation d'un problème de séparation non linéaire dans l'espace de représentation en un problème de séparation linéaire dans un espace de re-description de plus grande dimension. Cette transformation est réalisée *via* une fonction **noyau**. En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur de SVM d'effectuer des tests pour déterminer celle qui convient le mieux pour son application.

2. Principe de fonctionnement d'un SVM

2.4 Cas non linéaire

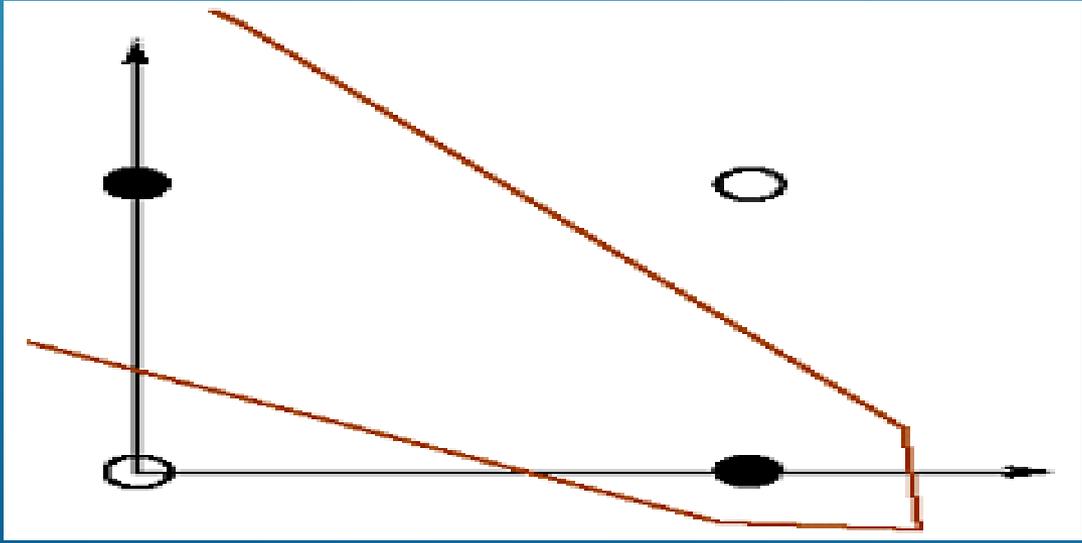
On peut citer les exemples de noyaux suivants : polynomiale, gaussien, sigmoïde et laplacien.



2. Principe de fonctionnement d'un SVM

2.5 Exemple de transformation de cas non linéaire le cas XOR

Le cas de XOR n'est pas linéairement séparable, si on place les points dans un plan a deux dimension, on obtient la figure suivante Coordonnees des points :
 $(0,0)$; $(0,1)$; $(1,0)$; $(1,1)$

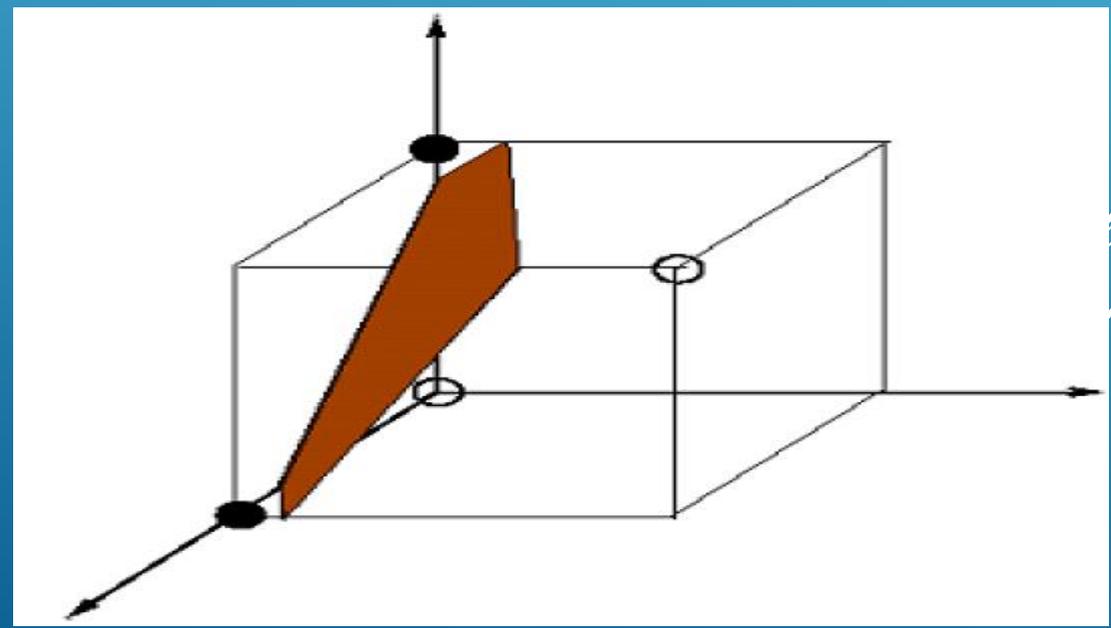


2. Principe de fonctionnement d'un SVM

2.5 Exemple de transformation de cas non linéaire le cas XOR

Si on prend une fonction polynomiale $(x,y) \rightarrow (x,y,x \cdot y)$ qui fait passer d'un espace de dimension 2 a un espace de dimension 3, on obtient un problème en trois dimensions linéairement separable :

- $(0,0) \rightarrow (0,0,0)$
- $(0,1) \rightarrow (0,1,0)$
- $(1,0) \rightarrow (1,0,0)$
- $(1,1) \rightarrow (1,1,1)$



3. Avantages et Inconvénient des SVM

3.1 Avantages :

SVM est une méthode de classification intéressante car le champ de ses applications est large, parmi ses avantages nous avons :

- ▶ Un grand taux de classification et de généralisation par rapport aux méthodes classiques.
- ▶ Elle nécessite moins d'effort pour designer l'architecture adéquate (petit nombre de paramètre à régler ou à estimer).
- ▶ La résolution du problème est donnée par des méthodes mathématiques .

3. Avantages et Inconvénient des SVM

3.2 Inconvénients :

- ▶ L'inconvénient majeur du classificateur SVM est qu'il est désigné ou conçu pour la classification binaire (la séparation entre deux classes)
- ▶ SVM ne convient pas aux grands ensembles de données en raison de son temps de formation élevé et il prend également plus de temps à s'entraîner que Naïve Bayes. Il fonctionne mal avec les classes qui se chevauchent et est également sensible au type de noyau utilisé.