

RAPPORT MACHINES A VECTEURS DE SUPPORT

SOMMAIRE

INTRODUCTION	1
VISUALISATION DES DONNÉES.	1
MACHINES A VECTEURS DE SUPPORT LINÉAIRES	1
ENTRAÎNEMENT DU MODÈLE	1
VISUALISATION DE LA FRONTIÈRE DE DÉCISION	2
MACHINES A VECTEURS DE SUPPORT GAUSSIEN (RBF)	2
ENTRAÎNEMENT DU MODÈLE	2
VISUALISATION DE LA FRONTIÈRE DE DÉCISION	3
MACHINES A VECTEURS DE SUPPORT POLYNOMIAL	4
ENTRAÎNEMENT DU MODÈLE	4
VISUALISATION DE LA FRONTIÈRE DE DÉCISION	5
MODÈLE CHOISIE	5

INTRODUCTION

Les machines à vecteurs de support (en anglais SVM) sont des algorithmes d'apprentissage automatique supervisé, dont le but est la séparation de l'ensemble de données par une frontière de façon à maximiser sa distance avec les données, les points les plus proches de la frontière sont appelés vecteurs support.

VISUALISATION DES DONNÉES.

Tout d'abord nous allons commencer par visualiser nos données afin d'avoir une vision globale de notre système.

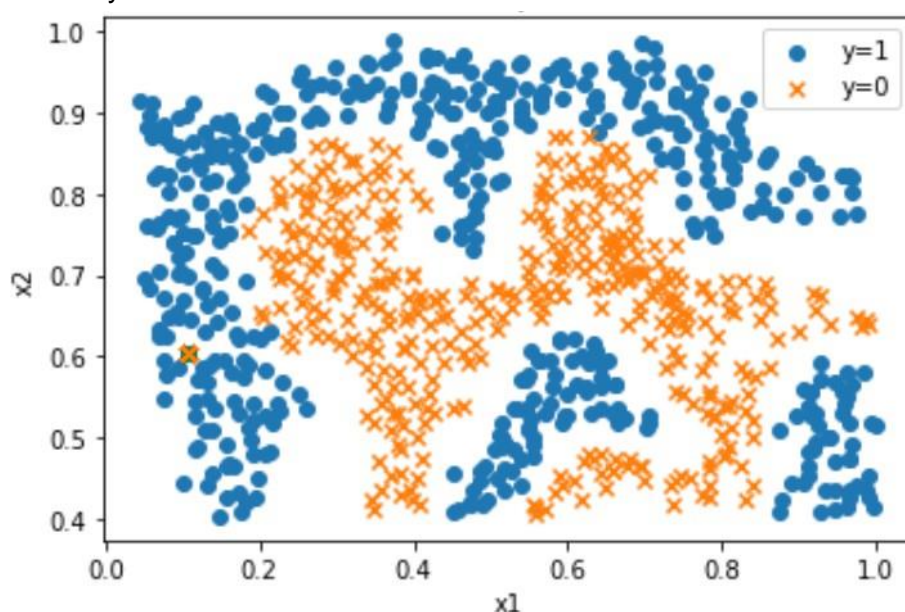


Figure 1: Visualisation de Données

Nous remarquons bien que nos données sont éparpillées un peu partout, ce qui veut dire qu'elles ne sont pas linéairement séparables.

MACHINES A VECTEURS DE SUPPORT LINÉAIRES

ENTRAINEMENT DU MODÈLE

Commençons par entraîner notre MVS linéaire.

```
svc = svm.SVC(kernel='linear', C=C).fit(X, y)
```

- kernel sa valeur représente le type du noyau utilisé (linear, rbf, poly).

- C représente le paramètre de régularisation, tel que plus sa valeur augmente plus notre modèle devient sensible aux valeurs aberrantes.

VISUALISATION DE LA FRONTIÈRE DE DÉCISION

Il est temps maintenant de visualiser la ligne séparatrice apprise par notre modèle linéaire, en variant à chaque fois le paramètre C .

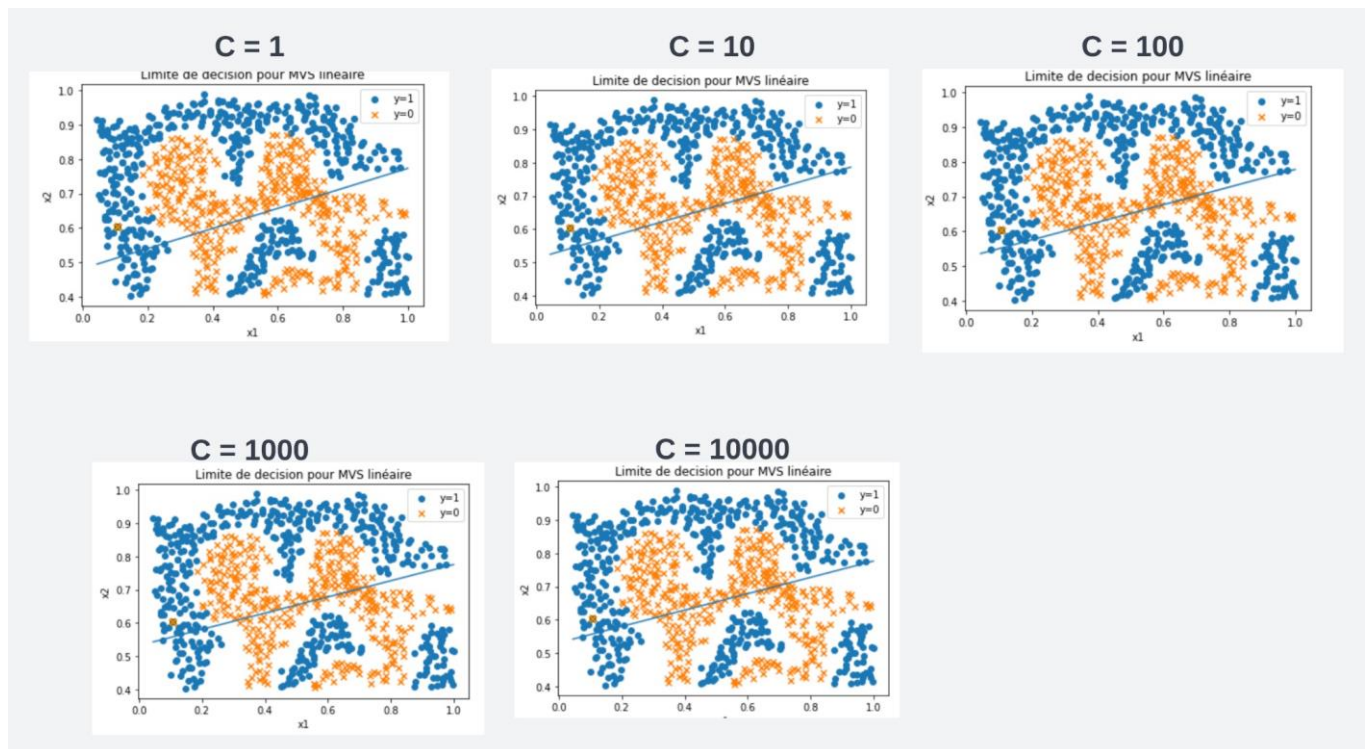


Figure 2: Frontière de décision MVS linéaire pour différentes valeurs de C

Comme nous pouvons le constater la ligne de décision n'a pas trop évolué à chaque fois que l'on a augmenté la valeur de C , et étant donné que nos données sont non linéairement séparables alors nous pouvons dire que le modèle linéaire n'est pas adapté à notre problème.

MACHINES A VECTEURS DE SUPPORT GAUSSIEN (RBF)

ENTRAINEMENT DU MODÈLE

Commençons par entraîner notre MVS gaussien (RBF).

```
rbf_svc = svm.SVC(kernel='rbf', gamma=0.7, C=C).fit(X, y)
```

- kernel sa valeur représente le type du noyau utilisé (linear, rbf, poly).
- C représente le paramètre de régularisation, tel que plus sa valeur augmente plus notre modèle devient sensible aux valeurs aberrantes.

VISUALISATION DE LA FRONTIÈRE DE DÉCISION

Nous allons commencer par visualiser la frontière de décision en variant à chaque fois la valeur de C sans toucher à la valeur de gamma, et puis celle de gamma avec C constant.

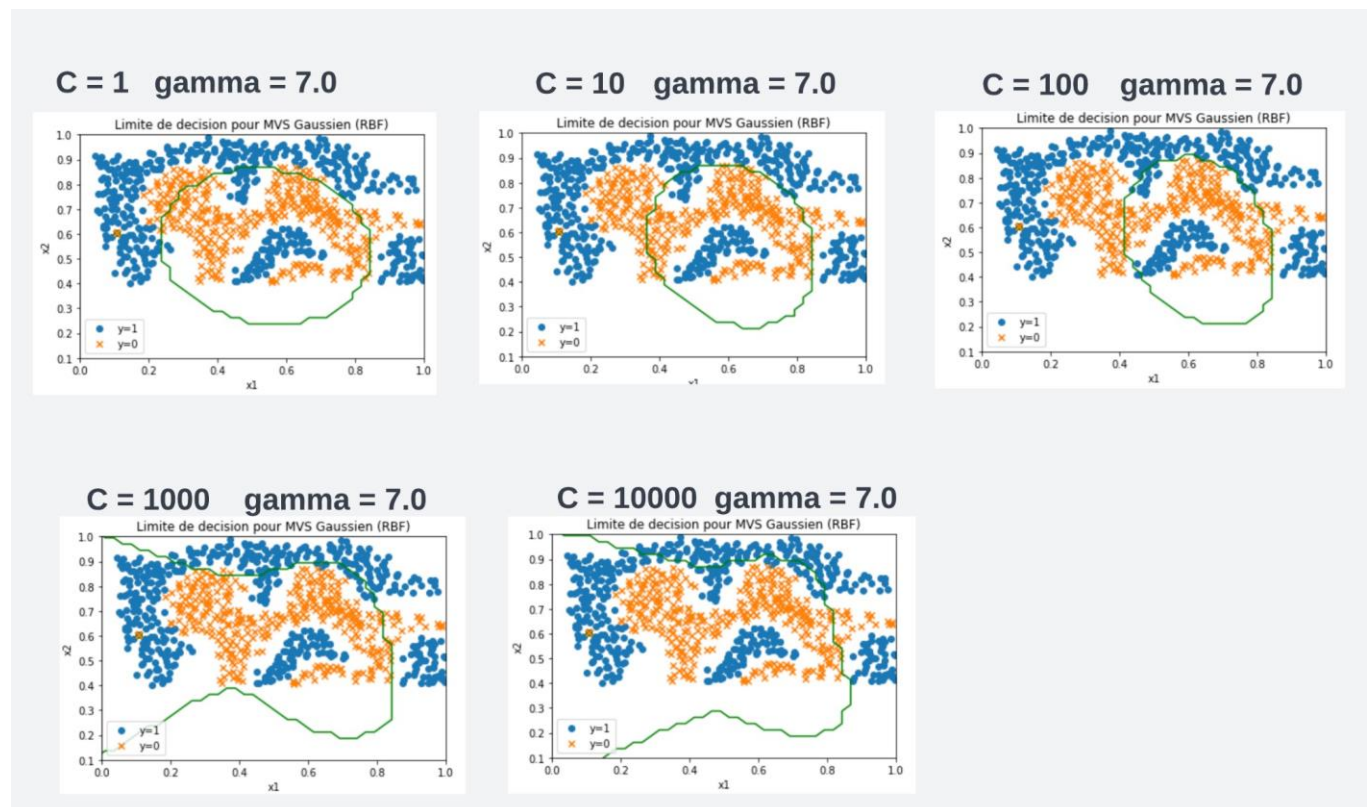


Figure 3: Frontière de décision pour MVS gaussien (C variant, gamma constant)

La frontière de décision ne semble pas vraiment faire du bon travail à chaque fois que l'on augmente la valeur de C.

Essayons maintenant d'augmenter la valeur de gamma avec C constant.

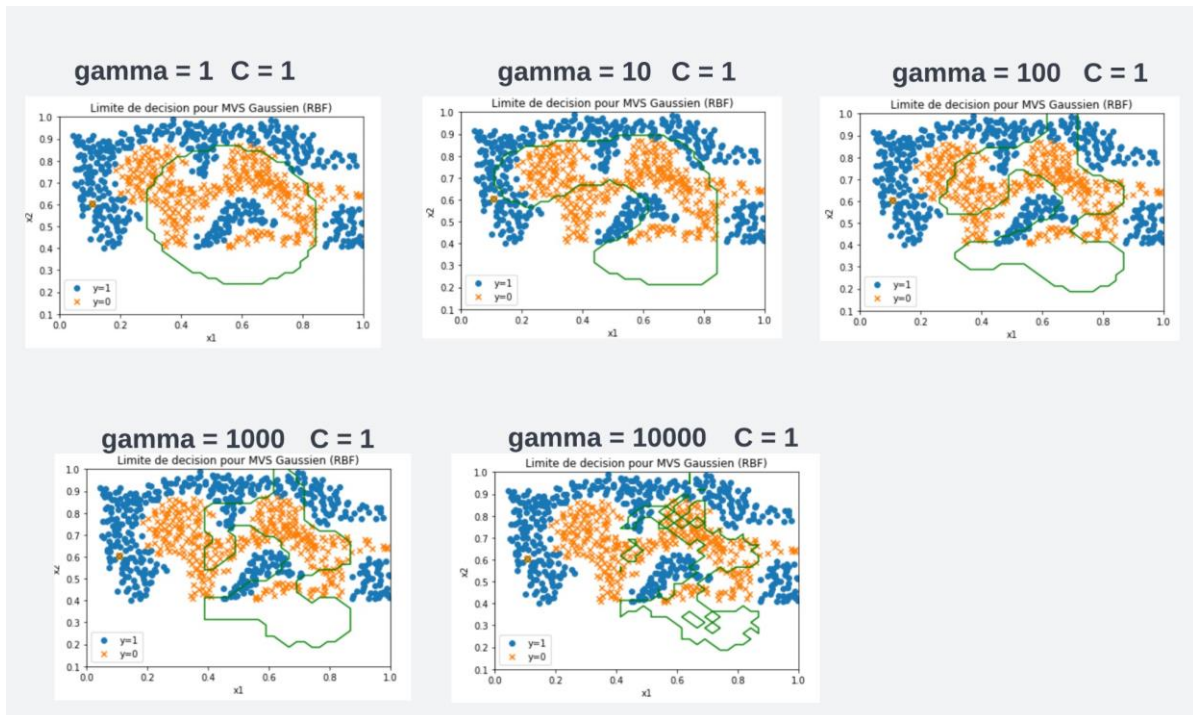


Figure 4: Frontière de décision pour MVS gaussien (C constant, gamma variant)

En analysant le figure du dessus on constate les points suivants:

- gamma = 1 : notre modèle ne semble pas avoir bien appris étant donné que la limite de décision ne sépare pas bien nos données.
- gamma = 10 : notre frontière de décision commence à bien rentrer dans le nuage de points, le modèle commence à apprendre, **on pourrait même se demander si ça ne serait pas les bons paramètres à choisir pour notre problème.**
- gamma > 100 : le modèle est en overfitting.

MACHINES A VECTEURS DE SUPPORT POLYNOMIAL

ENTRAINEMENT DU MODÈLE

Commençons par entraîner notre MVS polynomial.

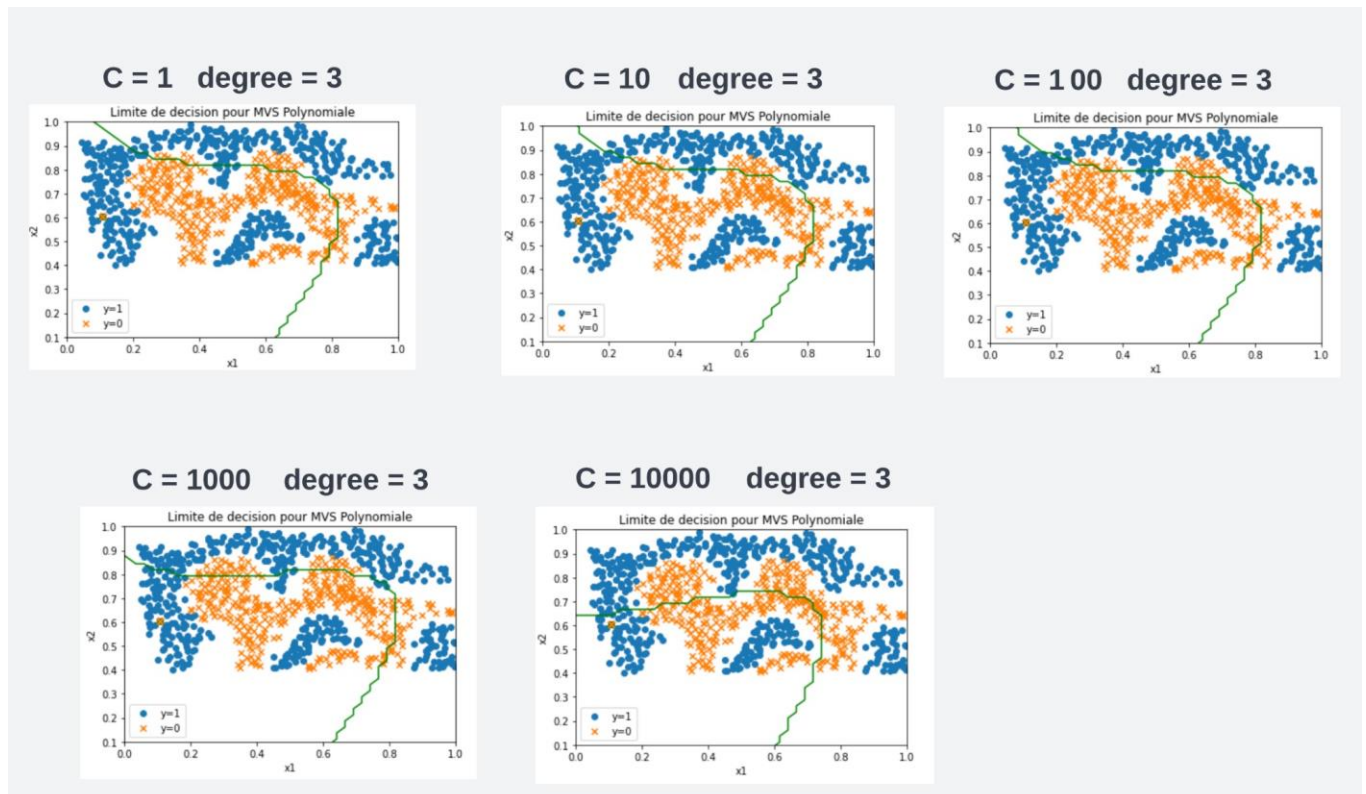
```
poly_svc = svm.SVC(kernel='poly', degree=3, C=C).fit(X, y)
```

- kernel sa valeur représente le type du noyau utilisé (linear, rbf, poly).

- C représente le paramètre de régularisation, tel que plus sa valeur augmente plus notre modèle devient sensible aux valeurs aberrantes.
- degree n'est autre que le degré du polynôme.

VISUALISATION DE LA FRONTIÈRE DE DÉCISION

Nous allons commencer par visualiser la frontière de décision en variant à chaque fois la valeur de C sans toucher au degré.



Comme nous pouvons le voir, la frontière de décision ne semble pas trop changer à chaque fois que l'on varie la valeur de C , en plus de celà elle sépare trop mal nos données.

MODÈLE CHOISIE

Le type de VMS choisie est le gaussien avec comme paramètres : $C = 1$, $\gamma = 10$ étant donné que c'est celui qui répond le plus à notre besoin.

La figure suivante nous montre la limite de décision de notre modèle, ainsi que les vecteurs de support.

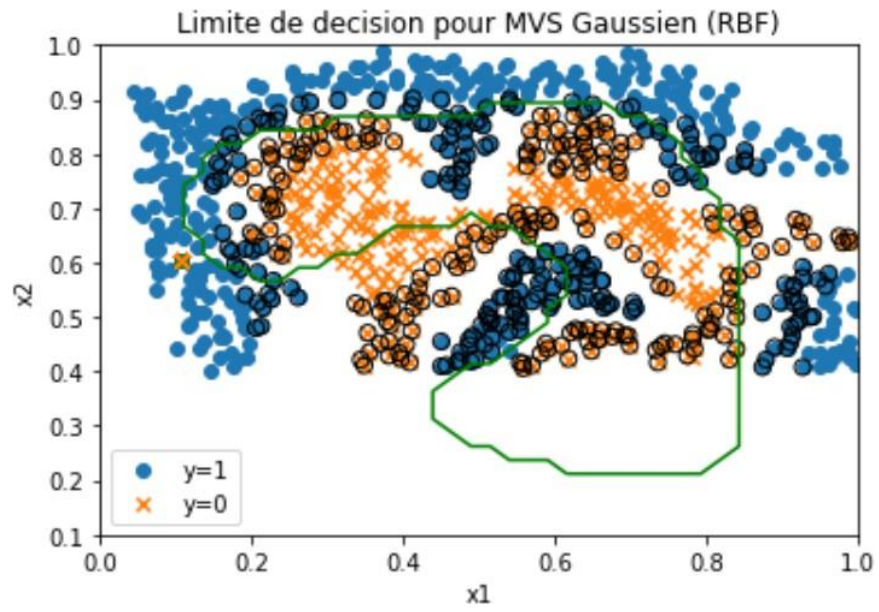


Figure 5: Limite de décision et vecteurs de support MVS Gaussien