TP/TD Optimisation Classique 2020 EMSE

- Majeure Science des Données

Séparateurs à Vaste Marge pour la classification (et la régression )

Yuteng WANG

Liwei XU

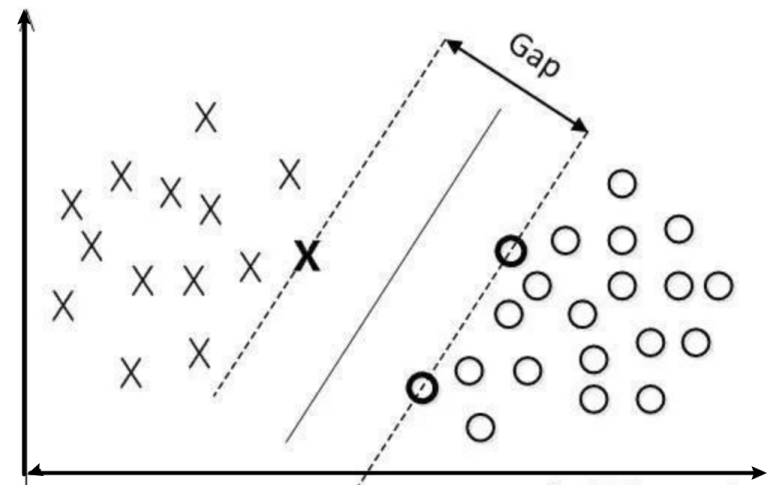
Wenxu ZHAO



SVM

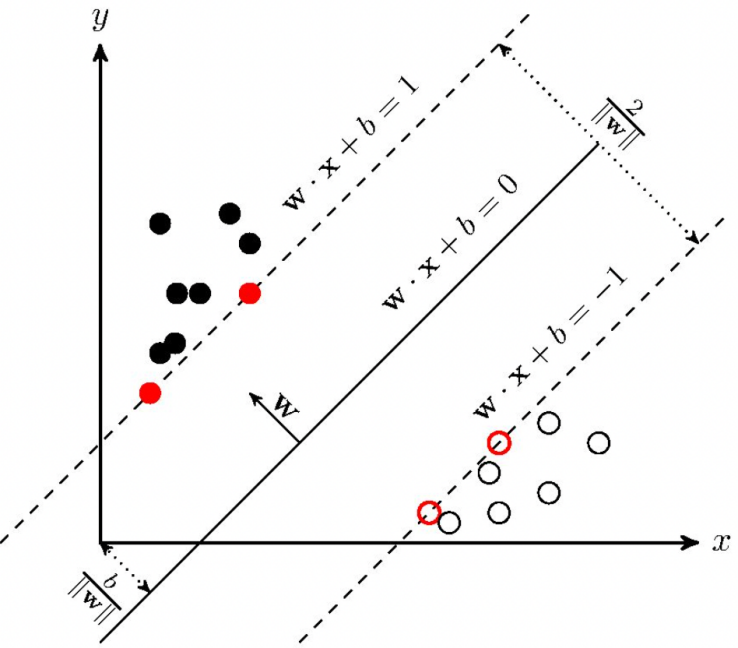
Comme le classificateur logistique, le MVC est un modèle de classification à deux classes dont le modèle de base est défini comme un classificateur linéaire avec le plus grand espacement sur l'espace des caractéristiques, dont la stratégie d'apprentissage consiste à maximiser l'espacement, ce qui se traduit en fin de compte par la solution d'un problème de planification quadratique convexe.

Pour l'ensemble de données suivant, il y a deux classes représentées par x et ○ respectivement. Ensuite, nous voulons trouver une courbe pour distinguer ces deux classes. Comme vous pouvez l'imaginer, il existe une infinité de courbes de ce type, comment trouver la courbe optimale ? Et c'est pour trouver la courbe qui a le plus grand intervalle (Gap), et ensuite nous pouvons nous classer en fonction de cette courbe. Et cette courbe, nous l'appelons l'hyperplan. Le x en gras et ○ dans la figure, nous appelons ces vecteurs des vecteurs de soutien.



Et en fait, nous rencontrons rarement la situation décrite ci-dessus où nous pouvons classer les données directement en utilisant une ligne droite, et le plus souvent, les données ne peuvent pas être classées par une ligne droite. Par exemple, la situation suivante

est l'hyperplan de séparation, et il existe un nombre infini de ces hyperplans pour un ensemble de données séparables linéairement (c'est-à-dire des machines perceptuelles), mais l'hyperplan de séparation avec la plus grande séparation géométrique est le seul

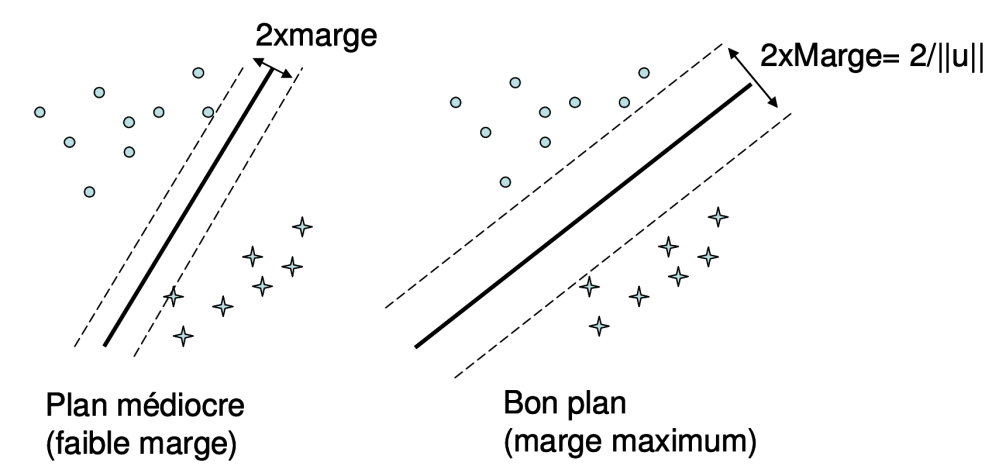


Ce critère correspond à la marge entre les deux classes : un hyperplan sera

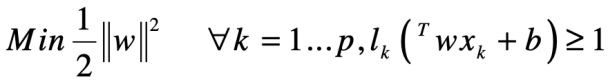
considéré comme meilleur dès lors qui sépare au mieux les deux classes.

Cette marge se définie de la manière suivante : M = 1/ ||w||

D’où la nécessité de minimiser ||w|| afin de maximiser la marge.

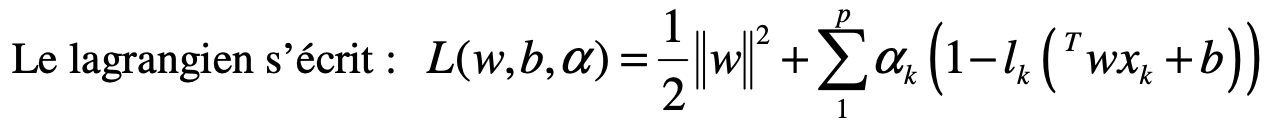


Dans le premier cas la marge sépare de manière correcte les deux classes, mais semble faible et améliorable, dans le second cas, les deux classes sont séparées par un hyperplan qui maximise la distance entre les points (de support) de chaque classe. Ces points, qui constituent la frontière entre appartenance à la classe et marge sont les points clés qui définissent l’hyperplan H



La théorie du modèle de SVM s’appuie sur le Lagrangien lié à cette maximisation,

On définit le lagrangien :

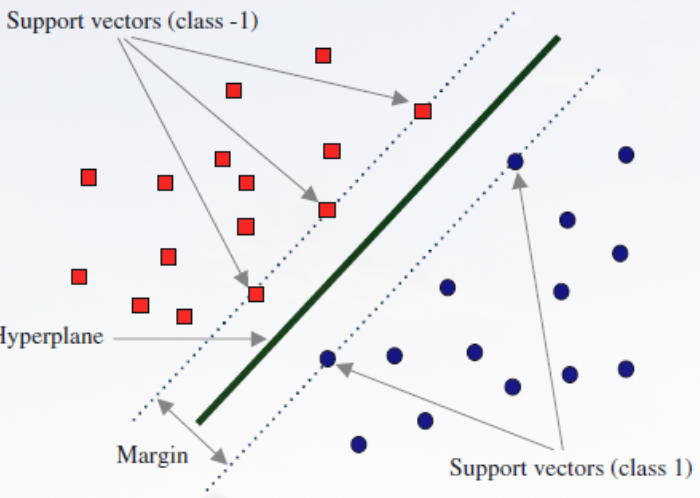


En finalité on peut estimer les inconnues primales w et b (après avoir estimé les inconnues duales introduites) et donc définir l’hyperplan H

Dans le cas linéairement différentiable, le point de données le plus proche de l'échantillon de l'ensemble de données d'entraînement par rapport à l'hyperplan de séparation est appelé vecteur de support, qui est le point qui rend les contraintes du graphe égales, c'est-à-dire qui satisfait à la

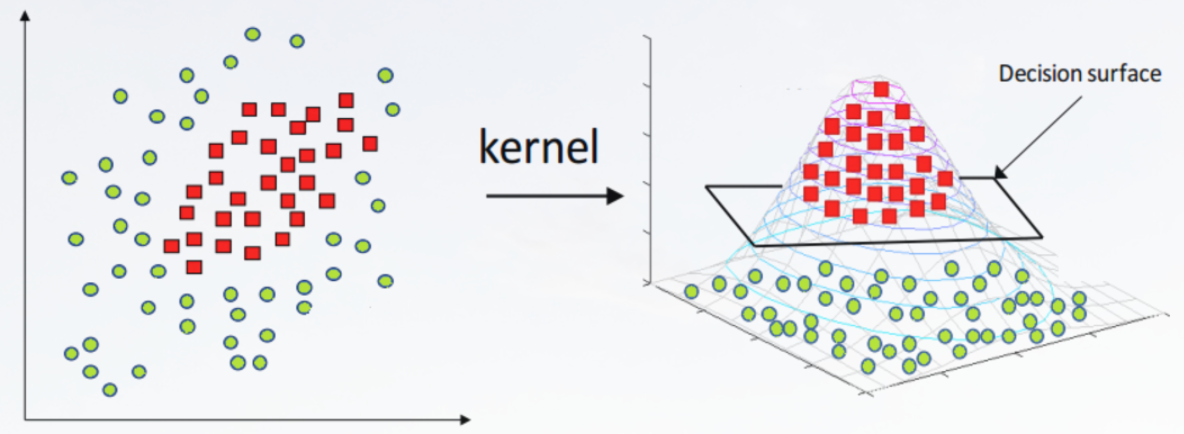


C'est-à-dire tous les points sur la ligne XTW+b=1 ou la ligne XTW+b=-1.



Enfin, et pour expliquer la présence de noyau dans la suite de la démarche de

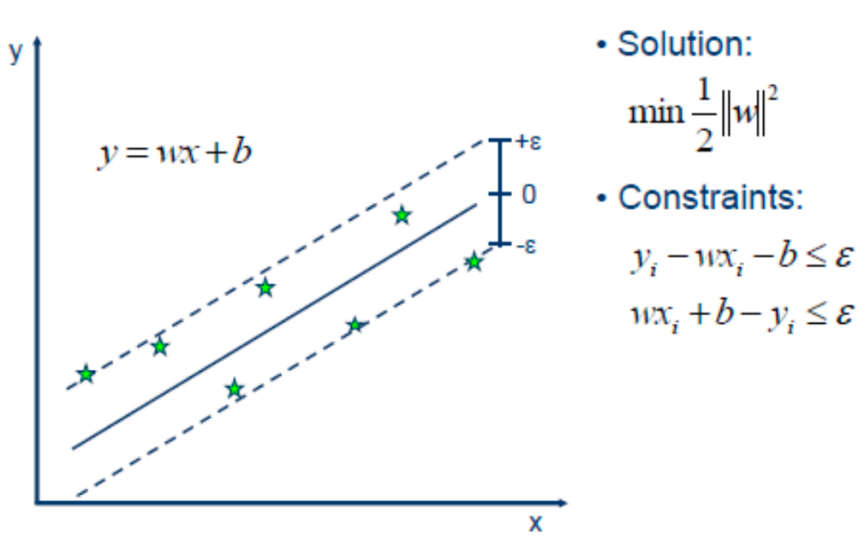
recherche il faut introduire les cas ou les données ne sont pas séparables .



Dans le premier cas un hyperplan existe afin de séparer les données, dans le second cas on passe par une fonction auxiliaire, le noyau, afin de pouvoir trouver un espace (de dimension plus grande souvent) dans lequel les données sont séparables.

**svr**

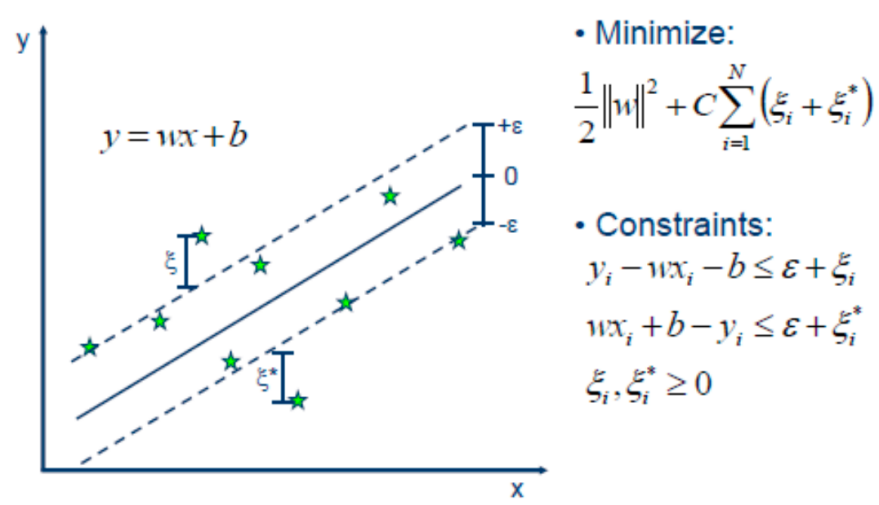
Nous savons que le modèle de régression linéaire le plus simple consiste à trouver une courbe qui minimise les résidus. De même, le SVR consiste à trouver un hyperplan qui minimise la distance entre toutes les données et cet hyperplan.seule la conclusion va être différente et les contraintes que l’on fixe lors de la maximisation



Cependant, dans une volonté de trouver le meilleur prédicteur, nous allons

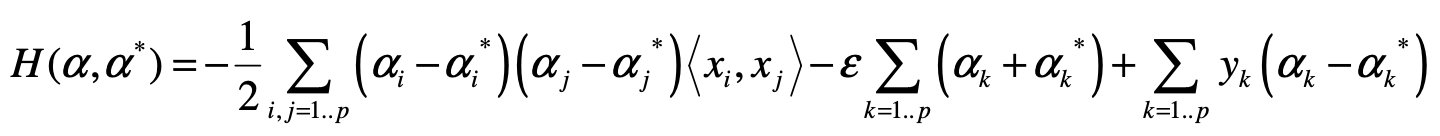
également considérer les cas ou l’on aurait des données « absurdes » ou très

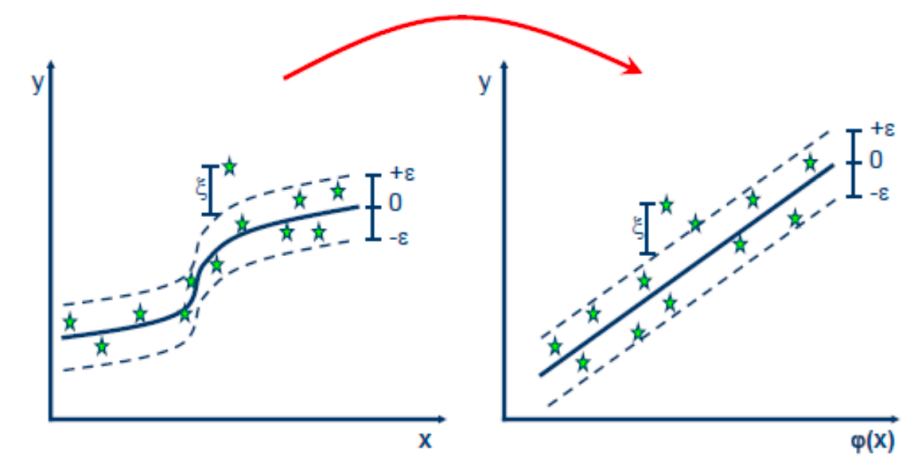
éloignées du reste du jeu, et cela en introduisant un coefficient d’erreur :



Le SVR est une application du MVC, et l'idée de base est la même, à quelques différences près. En utilisant le SVR pour l'analyse de régression, comme avec le SVM, nous devons trouver un hyperplan, la différence est que dans le SVM nous voulons trouver un hyperplan avec le plus grand espacement (gap), alors que dans le SVR nous définissons un ε, comme montré ci-dessus, qui définit les résidus des points de données dans la région à l'intérieur de la ligne en pointillés comme 0, et la distance des points de données à l'extérieur de la région en pointillés (vecteurs de support) à la limite de la ligne en pointillés comme les résidus (ζ). Comme pour le modèle linéaire, nous voulons que ces résidus (ζ) soient minimaux. En gros, le SVR consiste à trouver une région de bande optimale (2ε width) et à régresser les points en dehors de la région

Comme vous le verrez et devez le comprendre (refaites les calculs), la fonction duale s’écrit :





Pour les modèles non linéaires, la fonction du noyau est mise en correspondance avec l'espace des caractéristiques en utilisant la même fonction du noyau que le MVC, suivie d'une régression

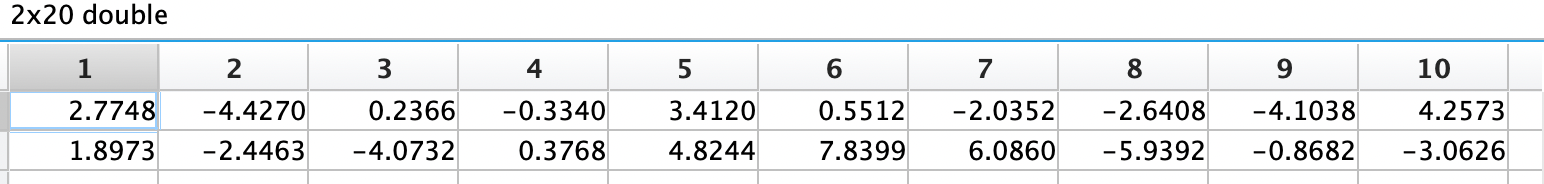
load antennes\_2d\_train  C  S  % Dans le fichier antenne,

C contient X' et % S contient Y'

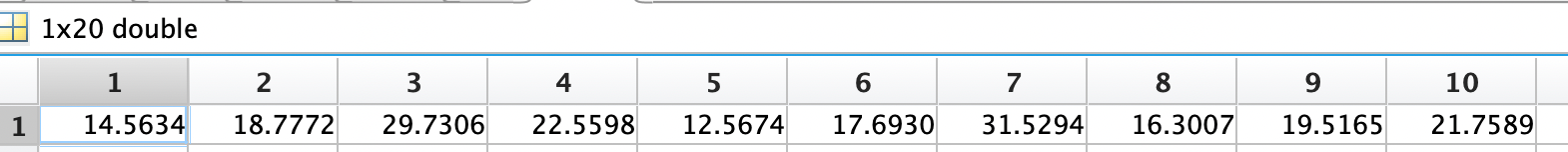
X=C';

Y=S';

X:



Y:



Pas 1 : Noyau Gaussien

Ici, on choit = 10 ;

Pas 2 : boucle de gradient

, , = [B,-B;-B,B], 0.5\*, =[Y’ ;-Y’], =

Jusqu’à ,

Comme pour les SVM, il faut se servir des points supports.

Points supports sont les points : > . Seulement les points soient à l’extérieur d’une marge de taille 2 ont un effet sur le calcule de b.

Pas 3: calcul du b

Comme pour les SVM, Il faut se servir des points supports. On a identifié les points support, pour lesquels le multiplicateur de lagrange ou es nul et donc :

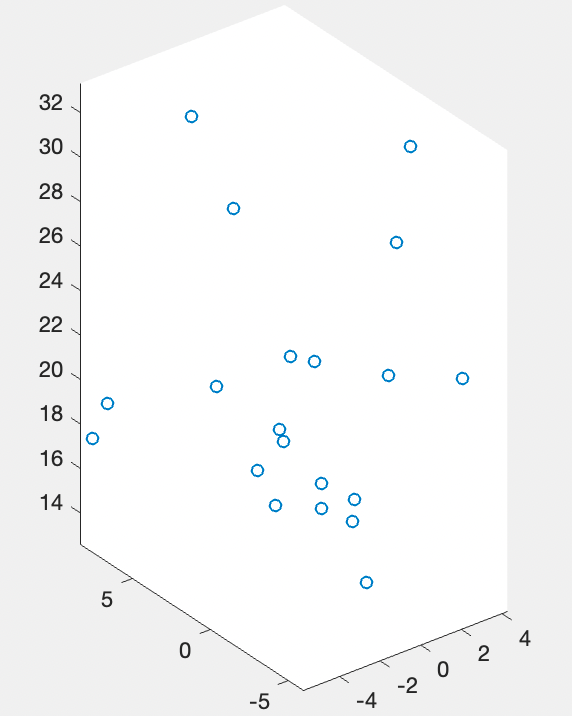
ou

On calculer le b pour chaque point support, puis on les moyens. On devrait faire le moyen pour éviter les erreurs numériques.

save Modele X Y alph b  % On stocke le materiel qui permetra de reconstruire

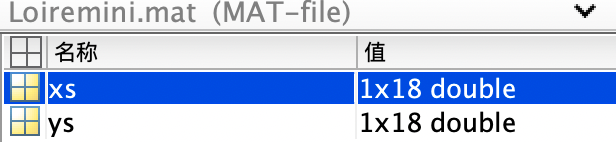
Dessin\_Modele (fonction de regression)

**Resultat:**

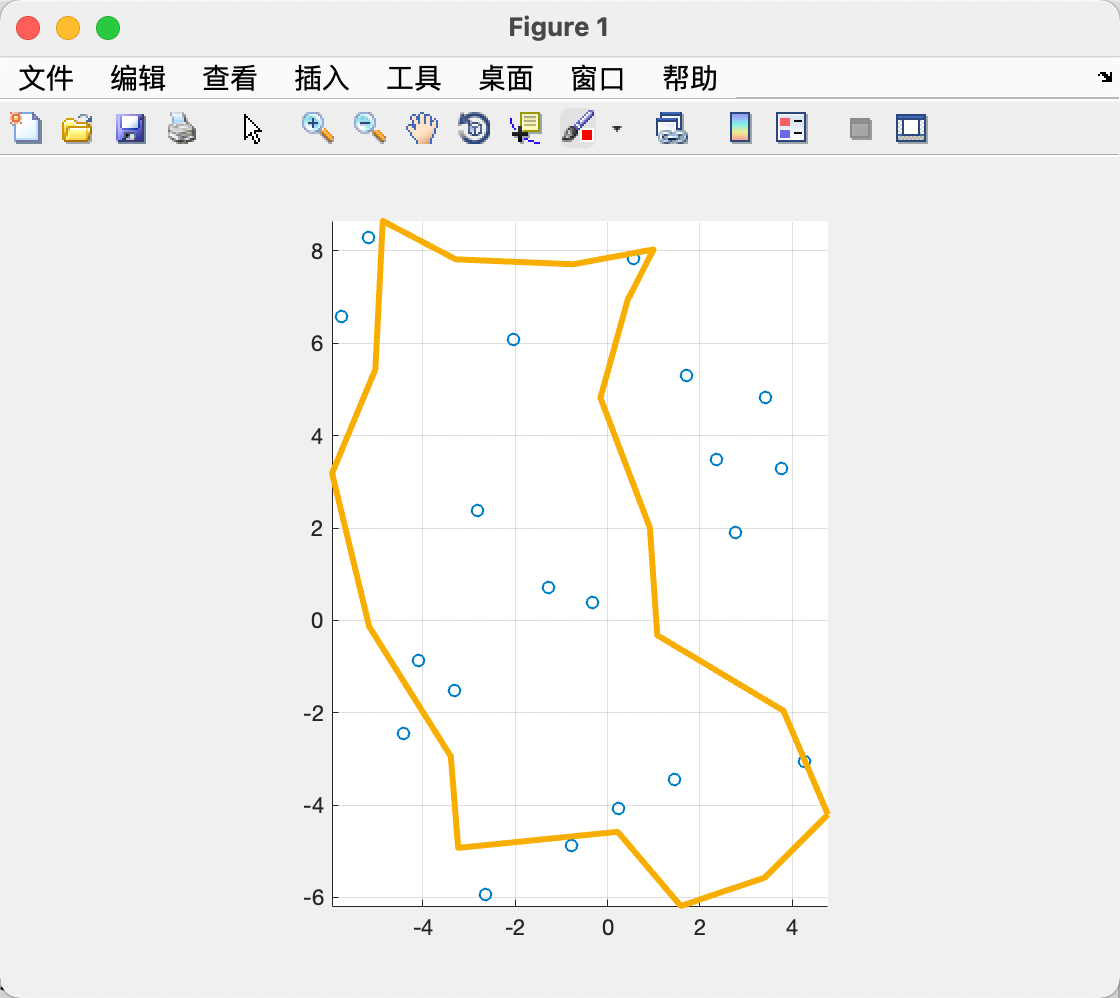


plot3(X(1,:),X(2,:),Y, 'o')

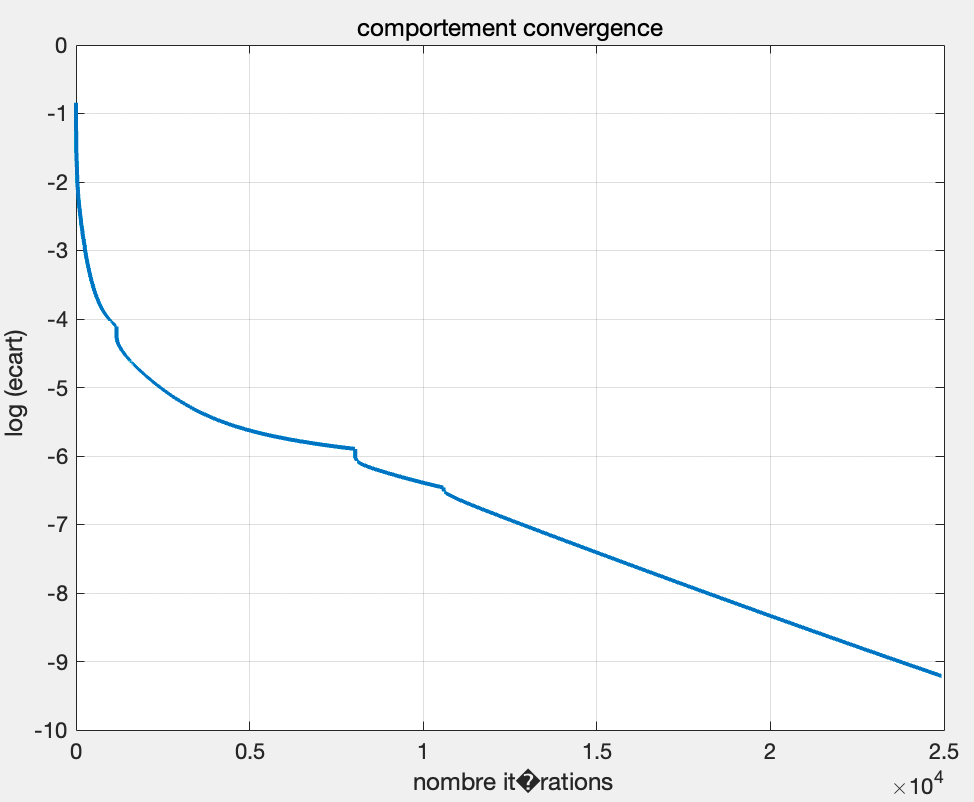
nb iterations: 24915 critere: 9.998581e-05



Loiremini



On dessine la Loire



dessin des performances de la convergence