



Projet Analyse de Données

Auteur :

AZDAD Bilal
MISSOURI Taha

Département Sciences du Numérique - Première année
2022-2023

Contents

1	Introduction	3
2	Extension 1 : SVM à noyau gaussien	3
3	Extension 2 : optimisation du SVM à noyau gaussien	4
4	Extension 3 : SVM linéaire à marge souple	4
5	Extension 4 : SVM à noyau gaussien et marge souple	5
6	Extension 5 : classification de données en dimension 3	5
6.1	SVM à noyau gaussien à 3D	5
6.2	SSVM à marge souple et noyau gaussien en 3D	6
7	Conclusion	6

1 Introduction

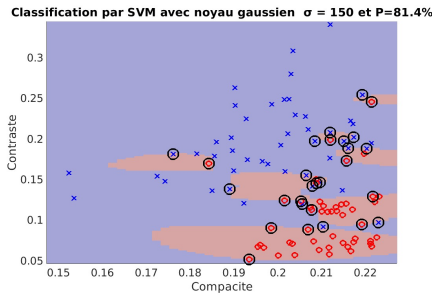
Le projet que nous abordons traite de la méthode de classification de données "Support Vector Machines" (SVM) dans sa formulation primale et duale. Nous étudions également différentes extensions de ce classifieur pour traiter des problèmes où les données suivent des structure plus complexes.

2 Extension 1 : SVM à noyau gaussien

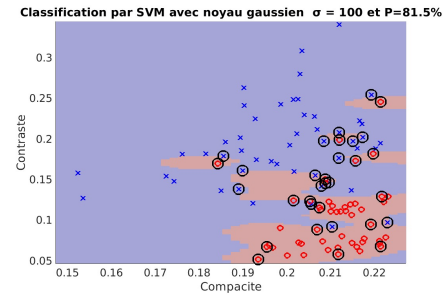
Il est rare que des données non filtrées soient linéairement séparables. Pour pallier ce problème, on peut appliquer aux points x_i une transformation non linéaire. Une technique souvent utilisée consiste à faire le "coup du noyau", on se propose d'appliquer un noyau Gaussien :

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

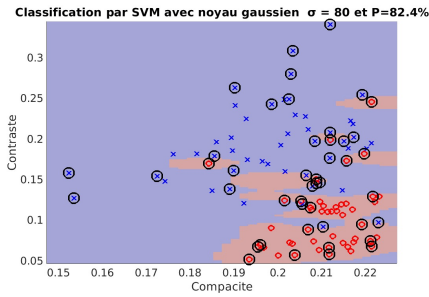
Ce terme apparaît dans la formulation duale et caractérise la transformation non linéaire par σ :



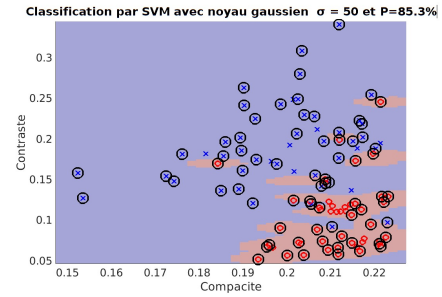
(a) Figure 1



(b) Figure 2



(c) Figure 3



(d) Figure 4

Figure 1: Classification par SVM noyau Gaussien pour différentes valeurs de sigma

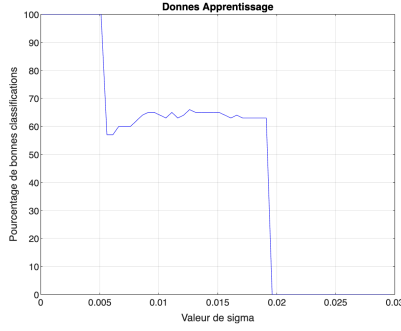
Observation :

Si sigma est petit, seuls les échantillons très proches de chaque point de données auront un poids significatif dans le calcul de K_{ij} . Cela peut conduire à des frontières de décision très complexes et à trop grande sensibilité aux données d'entraînement cela conduit à un surajustement, car le modèle peut être trop adapté aux données d'entraînement spécifiques.

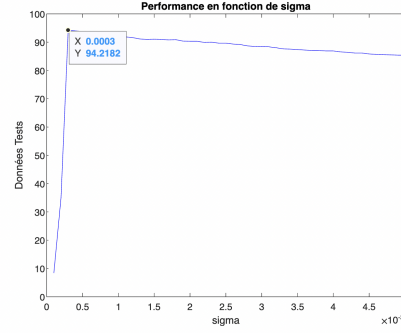
A l'inverse si sigma est grand, les échantillons plus éloignés auront une influence considérable dans le calcul de la fonction noyau. Cela peut conduire à des frontières de décision plus régulières. Cependant, une valeur trop grande de sigma peut entraîner une perte de détails fins dans les données et baisser la capacité de généralisation dans sa finesse.

3 Extension 2 : optimisation du SVM à noyau gaussien

On comprend donc que sigma est le paramètre à optimiser, les deux graphes suivant permettent d'obtenir σ_{opt}



(a) Données Apprentissage



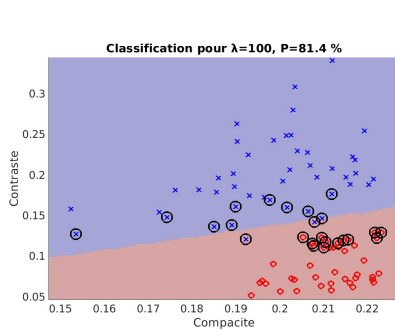
(b) Données Tests

Figure 2: Determination de σ_{opt}

la première courbe permet de réduire l'intervalle des sigma optimaux à l'aide des données d'apprentissage, la deuxième courbe qui prend en compte les données Test permet de trouver un résultat maximal de généralisation pour un pourcentage 94.2% pour $\sigma_{opt} = 0.0003$.

4 Extension 3 : SVM linéaire à marge souple

L'extension de la classification SVM avec la marge souple vise à classer des données non linéairement séparables en introduisant des variables de ressort pour tolérer certaines erreurs de classification de bases, ou données incorrectes à l'aide d'une valeur optimal du paramètre de régularisation . Cette extension permet d'améliorer la capacité de généralisation du classifieur SVM en rendant le modèle plus flexible et adapté aux données définissant des structures complexes.



(a) Classification lambda=100



(b) Détermination de lambda optimal

Figure 3: Utilisation de la fonction subspace_iter_3

La performance atteint un pic maximum de 85% pour $\lambda_{opt} = 17$

5 Extension 4 : SVM à noyau gaussien et marge souple

Nous avons combiné les concepts de la marge souple et du noyau gaussien pour obtenir une classification SVM hybride. L'objectif est d'obtenir une meilleure précision de classification en testant différentes valeurs de σ et de λ . Nous avons tracé une figure en 3D représentant le pourcentage de bonnes classifications en fonction de ces paramètres.

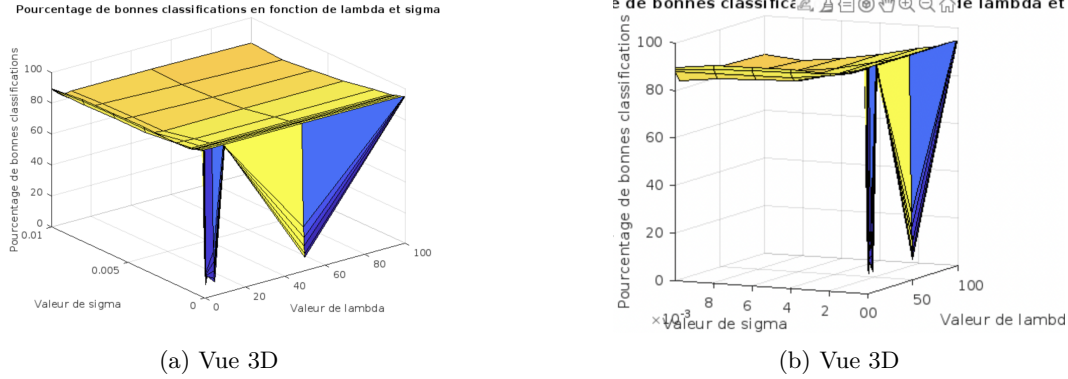


Figure 4: Determination du couple (lambda, sigma) optimal

Le point optimal qui offre le pourcentage maximal de 87% est $\sigma_{\text{opt}} = 0.0005$ et $\lambda_{\text{opt}} = 17$. Cela signifie que notre modèle SVM amélioré a réussi à classifier correctement la majorité des exemples avec une précision élevée.

6 Extension 5 : classification de données en dimension 3

Dans cette partie, nous traitons des données en 3D.

6.1 SVM à noyau gaussien à 3D

Nous avons étendue l'extension 1 à la dimension 3D, où nous avons appliqué la classification SVM à noyau gaussien sur un ensemble de données comportant trois caractéristiques : compacité, contraste et texture. Nous avons obtenu un taux de classification correcte de 78.6% pour les données de test.

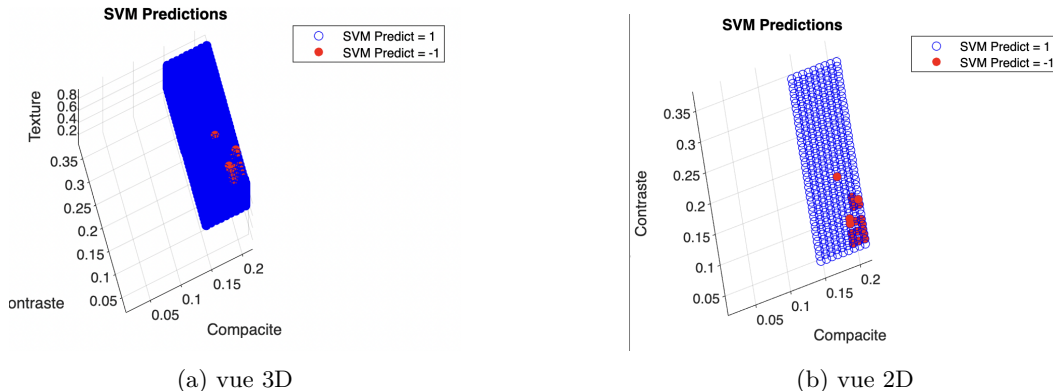


Figure 5: Classification SVM 3D noyau

6.2 SSVM à marge souple et noyau gaussien en 3D

Ici, nous avons combiné les approches à marge souple et à noyau gaussien afin d'optimiser notre modèle de classification dans l'objectif d'une meilleure généralisation. Les résultats obtenus ont démontré un taux plus élevé de classifications correctes, atteignant 88% .

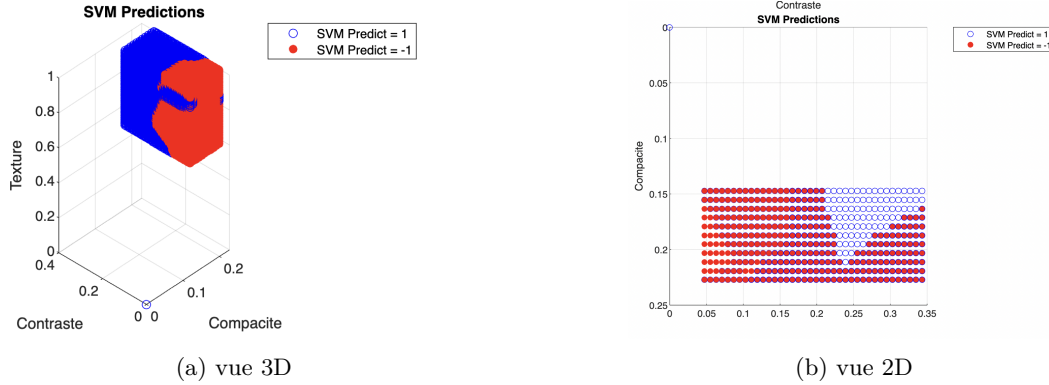


Figure 6: Classification SVM 3D noyau+marge souple

7 Conclusion

En conclusion, nous avons étudié le problème de classification à l'aide du SVM (Support Vector Machine). Nous avons examiné la formulation primale et la formulation duale du SVM linéaire, ainsi que ses extensions avec le noyau gaussien et l'utilisation de la marge souple toujours pour viser une meilleure généralisation.