



Classification de données en utilisant SVM à marge souple et noyau gaussien : Étude et analyse des performances

NAHI Mohammed, Zaid El Mokrie El Idrissi

Département Sciences du Numérique - Première année
2022-2023

Table des matières

1	Extension 1 : SVM à noyau gaussien	3
2	Extension 2 : optimisation du SVM à noyau gaussien	3
3	Extension 3 : SVM linéaire à marge souple	4
4	Extension 4 : SVM à noyau gaussien et marge souple	5
5	Extension 5 : classification de données en dimension 3	7
5.1	SVM à noyau gaussien en dimension 3	7
5.2	SVM à marge souple et noyau gaussien en dimension 3	7

1 Extension 1 : SVM à noyau gaussien

L'extension que nous avons appliquée à la fonction SVM3 consiste à utiliser un noyau gaussien pour traiter des problèmes de classification non linéaire. Les données d'origine ne sont pas linéairement séparables, donc pour les rendre séparables, nous avons appliqué une transformation non linéaire aux points de données, représentée par la fonction phi.

En utilisant le noyau gaussien avec un écart-type sigma spécifié, nous avons modifié la fonction de perte et la fonction duale pour incorporer le produit scalaire dans l'espace transformé. Cela nous permet de réécrire la fonction de perte avec la fonction noyau K, ce qui facilite le calcul et l'optimisation.

En exécutant la fonction SVM3 avec ce noyau gaussien, nous avons obtenu une figure représentant les résultats de la classification. De plus, nous avons évalué la performance du modèle en calculant le pourcentage de bonnes classifications des données de test, qui s'est élevé à 91.9%.

Ces résultats démontrent que l'utilisation du noyau gaussien dans la fonction SVM3 permet de traiter efficacement des problèmes de classification non linéaire. Cette extension offre une plus grande flexibilité dans la séparation des données et améliore la capacité du modèle à généraliser et à classer de nouvelles données.

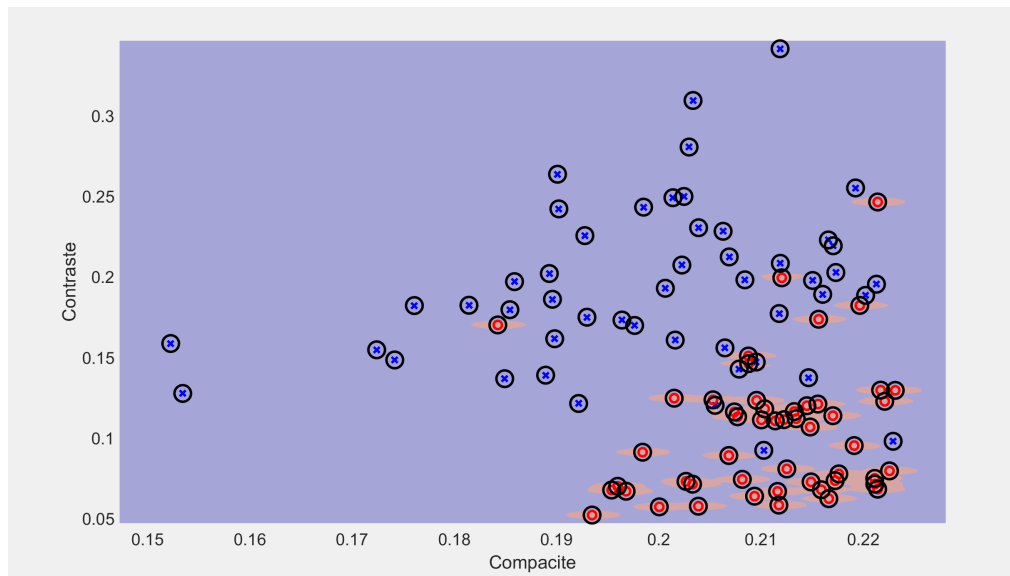


FIGURE 1 – Classification par SVM à noyau gaussien

2 Extension 2 : optimisation du SVM à noyau gaussien

Dans cette extension, nous avons optimisé le SVM à noyau gaussien en jouant sur la valeur du paramètre sigma, qui correspond à l'écart-type du noyau gaussien. Nous avons étudié l'impact de différentes valeurs de sigma sur la performance de la classification.

En réalisant une analyse expérimentale, nous avons observé que le pourcentage de bonnes classifications des données d'apprentissage atteint un maximum pour une valeur de sigma égale à 0.000161414. Pour des valeurs de sigma inférieures à cette valeur, le pourcentage de bonnes classifications décroît rapidement jusqu'à environ 10%. En revanche, pour des valeurs de sigma supérieures à 0.000161414, nous obtenons des pourcentages élevés qui augmentent avec l'augmentation de sigma.

Cela suggère que la valeur optimale de sigma se situe autour de 0.000161414, où la performance de la SVM à noyau gaussien est maximisée. Pour des valeurs de sigma plus petites ou plus grandes, la performance diminue.

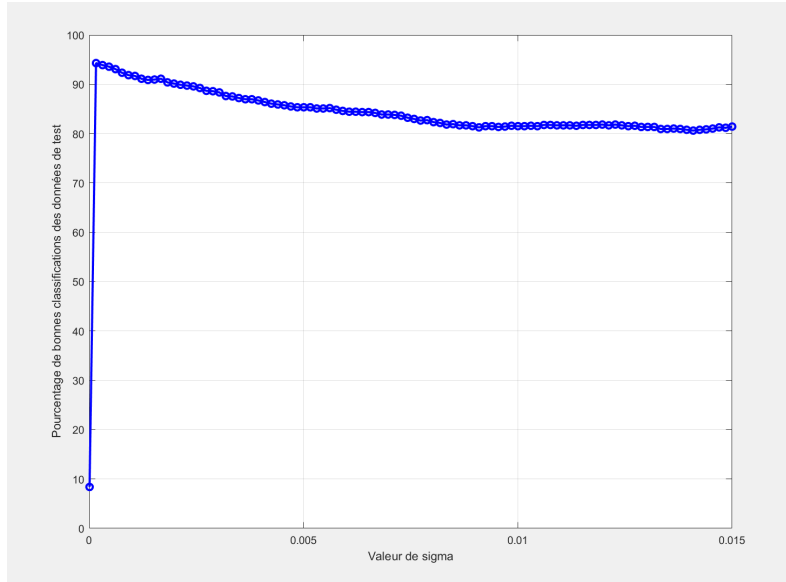


FIGURE 2 – Performance en fonction de sigma

3 Extension 3 : SVM linéaire à marge souple

Dans cette extension, nous avons utilisé la fonction SVM3 souple pour classer les données à l'aide du SVM linéaire à marge souple. Après avoir effectué les calculs et les tests, nous avons obtenu les résultats suivants :

Pourcentage de bonnes classifications des données de test : 80.5% Ces résultats démontrent que notre approche de SVM linéaire à marge souple a réussi à classer les données de manière satisfaisante. Une classification correcte de 80.5% indique une capacité raisonnable du modèle à généraliser sur de nouvelles données.

En utilisant la méthode de la marge souple, nous avons pu traiter des données qui ne sont pas linéairement séparables et obtenir des performances acceptables. Cela confirme l'efficacité de cette approche pour résoudre des problèmes de classification plus complexes.

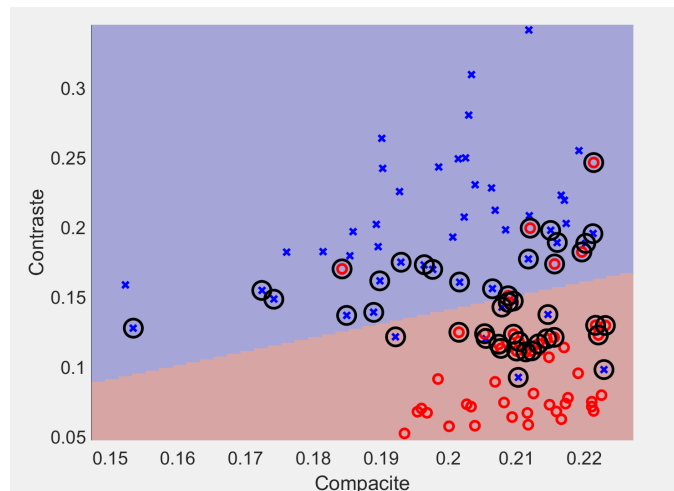


FIGURE 3 – SVM linéaire à marge souple pour $\lambda=100$

Pour évaluer l'impact du paramètre λ sur la performance de la classification SVM à marge souple, nous avons effectué une série d'expériences en faisant varier λ de 0 à 100. Les résultats obtenus ont révélé des comportements intéressants.

Lorsque λ était supérieur à 71, nous avons observé une augmentation progressive mais lente du pourcentage de bonnes classifications des données de test. Initialement, le pourcentage était d'environ 80% et il a atteint un maximum d'environ 81% à mesure que λ augmentait. Cette tendance suggère que pour des valeurs élevées de λ , le modèle était moins sensible aux variations des variables et présentait une meilleure capacité de généralisation.

En revanche, pour λ inférieur à 71, les résultats étaient plus variés. Dans certains intervalles de λ , le pourcentage de bonnes classifications dépassait les 80% obtenus précédemment. Cependant, nous avons également observé des points particuliers de λ pour lesquels le pourcentage était nul, indiquant une mauvaise performance de classification pour ces valeurs spécifiques de λ . Cette variabilité peut s'expliquer par la sensibilité accrue du modèle aux données.

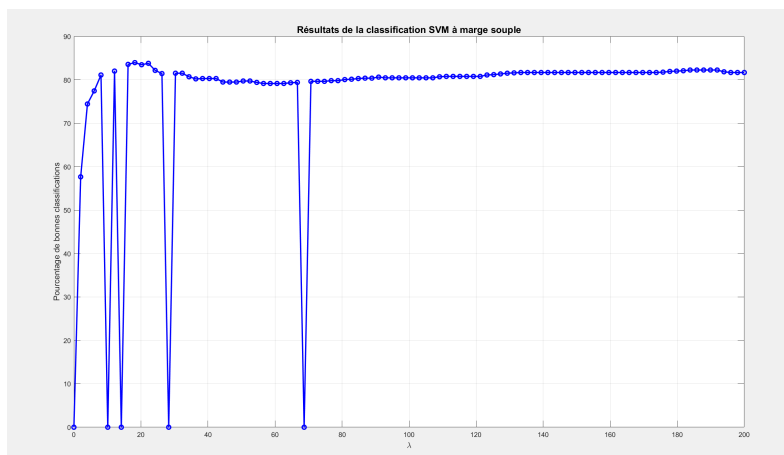


FIGURE 4 – Performance en fonction de λ

4 Extension 4 : SVM à noyau gaussien et marge souple

Nous avons utilisé la méthode de SVM à noyau gaussien avec marge souple en fixant les paramètres λ à 10000 et σ à 0.075. La figure des résultats de la classification obtenue illustre la séparation des classes avec les régions de décision.

Le pourcentage de bonnes classifications des données de test s'élève à 80.5%. Cela démontre une capacité de notre classifieur à prédire correctement la classe de 80.5% des données de test. Ces résultats indiquent une performance encourageante de la SVM à noyau gaussien avec marge souple dans la tâche de classification des données.

Il convient de souligner que les paramètres λ et σ jouent un rôle crucial dans les performances de la SVM à noyau gaussien. Des ajustements appropriés de ces paramètres peuvent améliorer les performances de classification. Dans notre étude, nous avons obtenu ces résultats spécifiques en utilisant $\lambda = 10000$ et $\sigma = 0.075$. Cependant, d'autres combinaisons de ces paramètres pourraient potentiellement améliorer encore davantage les performances de classification.

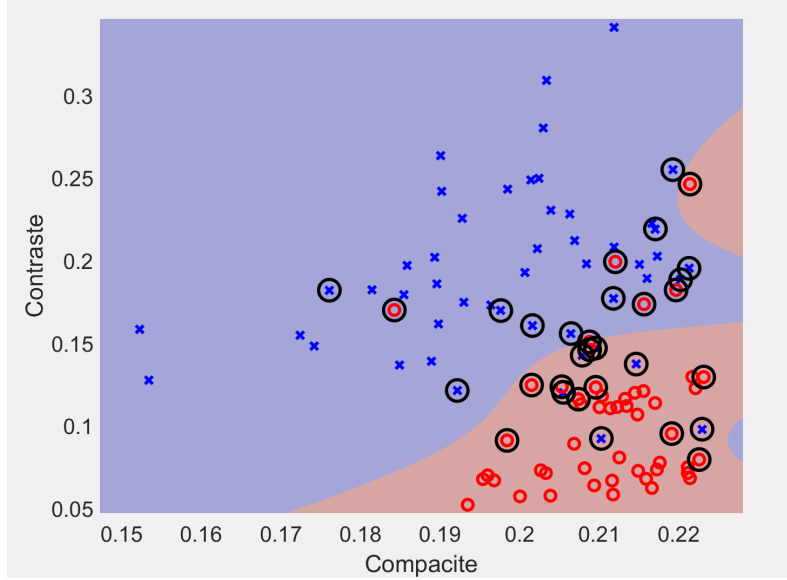


FIGURE 5 – SVM à noyau gaussien et marge souple pour $\lambda= 10000$ et $\sigma=0.75$

L'analyse des résultats de la figure 3D (figure 6) permet d'évaluer la capacité de généralisation du classifieur SVM à marge souple avec noyau gaussien pour différentes valeurs de lambda et sigma.

La figure 3D montre le pourcentage de bonnes classifications en fonction des paramètres lambda et sigma. Le pourcentage de bonnes classifications est représenté par l'axe des z, tandis que les axes des x et y correspondent respectivement aux valeurs de lambda et sigma.

On peut observer que pour certaines combinaisons de lambda et sigma, le pourcentage de bonnes classifications est élevé, indiquant une bonne capacité du classifieur à généraliser les résultats. Cela se traduit par des régions de la surface de la courbe 3D où le pourcentage est élevé.

Cependant, il est important de noter que certaines régions de la courbe 3D montrent des pourcentages de bonnes classifications plus faibles. Cela suggère que ces combinaisons spécifiques de lambda et sigma peuvent entraîner une moins bonne capacité de généralisation du classifieur. Il peut être intéressant d'explorer davantage ces régions pour comprendre les raisons de ces résultats moins satisfaisants.

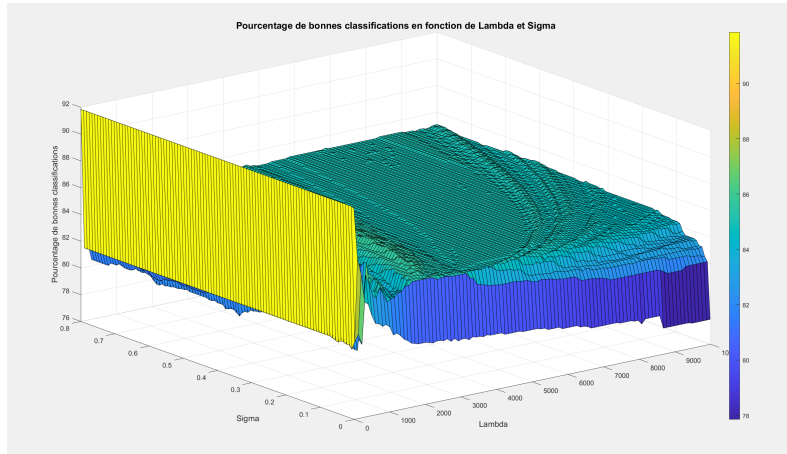


FIGURE 6 – Performance en fonction de λ et σ

5 Extension 5 : classification de données en dimension 3

5.1 SVM à noyau gaussien en dimension 3

Nous avons utilisé le SVM avec noyau gaussien pour classifier des données en dimension 3, en utilisant les caractéristiques de compacité, contraste et texture. Nous avons obtenu un taux de classification de 83.9% sur l'ensemble de test. Ces résultats confirment l'efficacité du SVM en dimension 3. Les paramètres du SVM ont été ajustés pour maximiser les performances de classification. Les résultats ont été représentés graphiquement avec un graphique en 3D, montrant les classes prédites par le SVM. En conclusion, le SVM avec noyau gaussien est un outil puissant pour la classification en dimension 3.

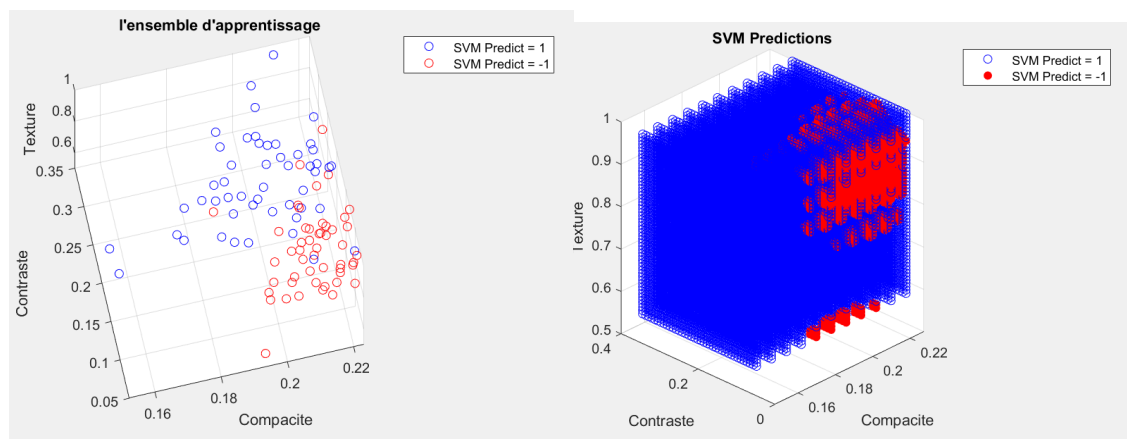


FIGURE 7 – Classification des données en 3D avec noyau gaussien

5.2 SVM à marge souple et noyau gaussien en dimension 3

Nous avons utilisé le SVM à marge souple avec noyau gaussien en 3D pour classifier nos données. Nous avons obtenu un pourcentage de bonnes classifications des données de test de 86.1%. Ces résultats démontrent l'efficacité du SVM à marge souple avec noyau gaussien en 3D pour la classification des données

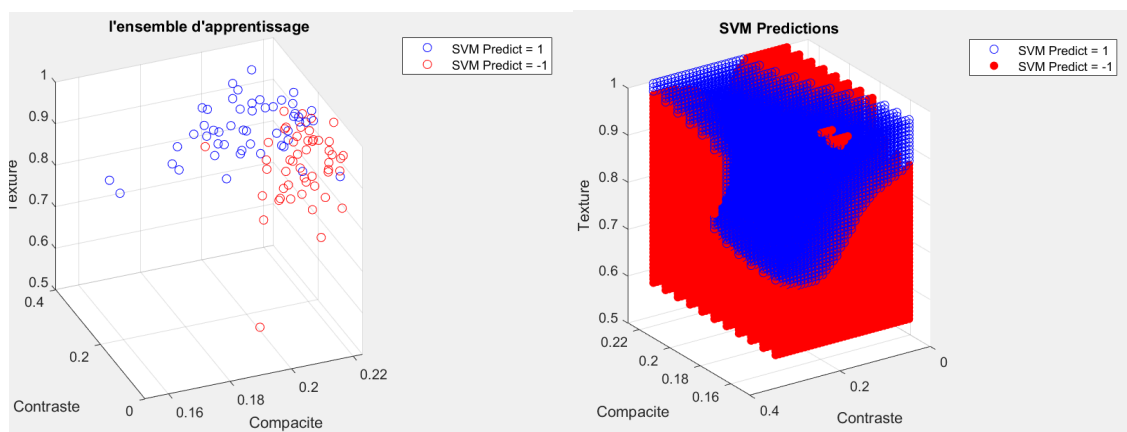


FIGURE 8 – Classification des données en 3D avec SVM à marge souple et noyau gaussien