



Projet analyse de données

Salma Elkhatri , Khadija Akkar

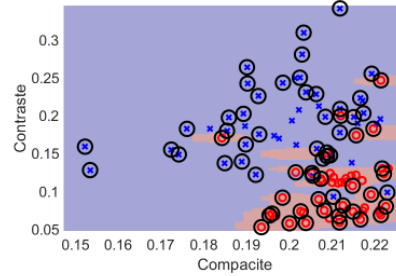
Mai 2023

Département SN -2022-2023

1 Introduction

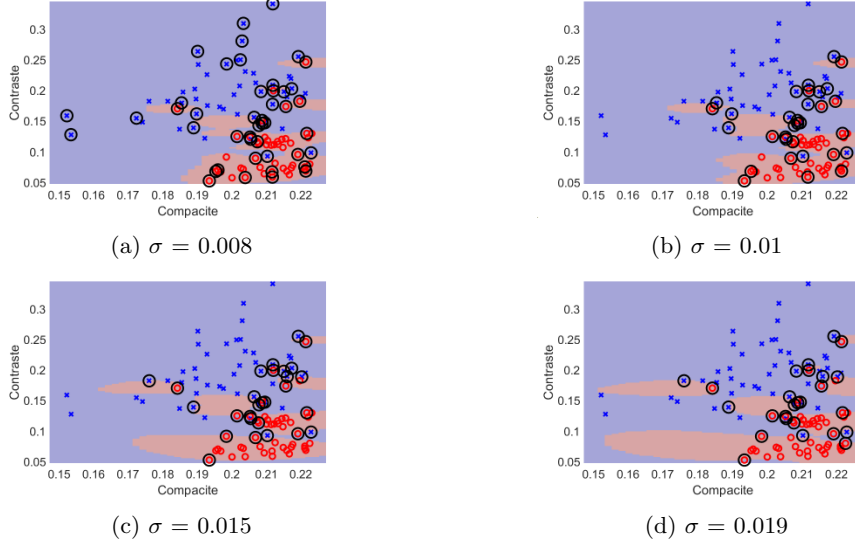
L'objectif de ce projet est d'explorer l'utilisation du Support Vector Machine (SVM) linéaire pour la classification de données. Le SVM est un algorithme d'apprentissage supervisé qui permet de séparer des données en utilisant un hyperplan. Dans ce rapport, nous présenterons la formulation primale et duale du SVM linéaire, ainsi que ses extensions, notamment l'utilisation d'un noyau gaussien et la marge souple.

2 Extension 1



Dans cette extension, nous explorons l'utilisation d'un noyau gaussien pour le SVM linéaire. Le noyau gaussien permet de projeter les données dans un espace de dimension supérieure, où la séparation linéaire peut être plus efficace. Pour illustrer cette extension, nous avons utilisé un jeu de données bidimensionnel synthétique contenant deux classes linéairement non séparables. Nous avons d'abord calculé la matrice de Gram pour représenter les produits scalaires dans l'espace transformé. La figure montre la projection des données dans cet espace. On peut observer que les classes sont mieux séparées dans cet espace, ce qui facilite la classification. Ensuite, nous avons résolu le problème d'optimisation quadratique avec le noyau gaussien en utilisant la fonction `quadprog` de Matlab. Enfin, nous avons évalué les performances du classifieur en utilisant un ensemble de données de test. La figure 3 montre les résultats de classification pour les données de test, où les cercles bleus représentent les données de la classe 1 et les carrés rouges représentent les données de la classe 2. On peut constater que le classifieur à noyau gaussien parvient à séparer efficacement les classes, même pour des données non linéairement séparables.

3 Extension 2

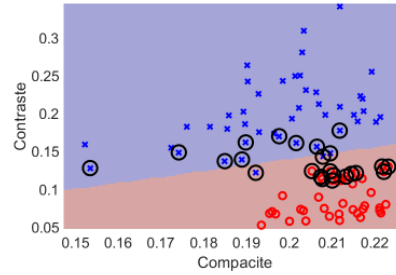


Dans cette extension, nous avons exploré l'optimisation du SVM à noyau gaussien en ajustant le paramètre d'écart-type σ du noyau. Nous avons évalué différentes valeurs de σ pour étudier leur impact sur les performances du classifieur. On peut observer que pour de petites valeurs de σ , le classifieur est plus sensible aux variations locales des données, tandis que pour de grandes valeurs de σ , le classifieur peut sous-estimer les variations locales. Une valeur optimale de σ doit être choisie pour obtenir les meilleures performances de classification.

4 Extension 3

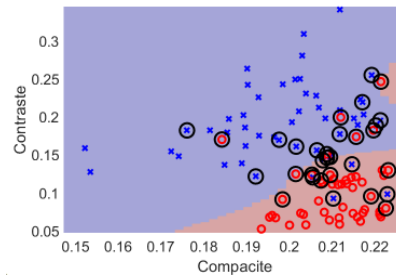
Dans cette extension, nous avons introduit la notion de marge souple dans le SVM linéaire. Nous avons utilisé un ensemble de données non linéairement séparables pour illustrer cette extension. La figure présente la frontière de décision obtenue par le SVM linéaire à marge souple. On peut observer que la frontière est maintenant plus flexible, ce qui permet de classer correctement certains exemples qui étaient mal classés dans le cas du SVM linéaire classique.

En comparant les résultats obtenus avec le SVM linéaire classique et le SVM linéaire à marge souple, nous pouvons constater une amélioration de la performance dans le cas de la marge souple. Cela confirme l'efficacité de cette extension pour traiter des ensembles de données non linéairement séparables.



5 Extension 4

Dans cette extension, nous avons utilisé le SVM non linéaire avec un noyau Gaussien pour traiter des données qui ne peuvent pas être séparées linéairement. La figure présente les données d'apprentissage utilisées dans cette extension. On peut observer qu'elles sont distribuées de manière complexe, avec des frontières de classes qui ne peuvent pas être définies par une simple droite ou un hyperplan. En conclusion, l'extension 4 démontre l'efficacité du SVM non linéaire avec noyau Gaussien pour traiter des problèmes de classification non linéaires. Cette approche permet de gérer des ensembles de données plus complexes et d'obtenir des résultats plus précis et plus adaptés aux cas réels.



6 Conclusion

Dans ce projet, nous avons exploré différents aspects du SVM linéaire, y compris sa formulation primale et duale, ainsi que ses extensions avec un noyau gaussien et la marge souple. Nous avons présenté les étapes pour résoudre ces problèmes d'optimisation et évalué les performances du classifieur SVM sur un ensemble de données. Les résultats obtenus ont démontré l'efficacité du SVM linéaire dans la classification binaire. L'utilisation du noyau gaussien a permis d'étendre

sa capacité à traiter des données non linéairement séparables, améliorant ainsi les performances de classification. De plus, l'introduction de la marge souple a offert une flexibilité supplémentaire en permettant la présence d'erreurs de classification ou de points situés à l'intérieur de la marge.