

INP ENSEEIHT

Institut national polytechnique en France

L'École nationale supérieure d'électrotechnique, d'électronique, d'informatique,
d'hydraulique et des télécommunications

Première année

Classification SVM

(Analyse de donnée - Projet)

Etude des différents extension de la classification SVM

Projet réalisé par *(Ayoub Bouchama)*
et *(Oussama Elguerraoui)*

Dirigée par *(Cuadrado Javier)*

Projet achevé et rendu le *(30 Mai)*

Lien utile vers les page des realisateurs du projet :

Ayoub Bouchama, *Département SN, Groupe F* ([Profil LinkedIn](#))

Oussama ElGuerraoui, *Département SN, Groupe F* ([Profil LinkedIn](#))

Table des matières

1	Introduction	3
2	Extension-1 : SVM à noyau gaussien	3
3	Extension-2 : Optimisation du SVM à noyau gaussien	3
4	Extension-3 : SVM lineaire à marge souple	4
5	Extension-4 : SVM à noyau gaussien et marge souple	5
6	Extension-5 : Classification de données en dimension 3	8
6.1	SVM à noyau gaussien en 3D	8
6.2	SVM à marge souple en 2D	8
6.3	SVM à marge souple et noyau gaussien en 3D	9
7	Conclusion	9

Table des figures

1	Résultats de la classification à l'aide de la fonction SVM3	3
2	Tracé du pourcentage de bonnes classifications en fonctions de σ	4
3	Résultats de la classification SVM à marge souple λ	5
4	Tracé du pourcentage de bonnes classifications en fonctions de λ	5
5	Tracé du pourcentage de bonnes classifications en fonctions du couple λ et σ	6
6	Tracé du pourcentage de bonnes classifications en fonctions du couple λ et σ	6
7	Tracé de λ_{max} en fonction de σ_{max}	7
8	Valeur des prédictions en fonction des paramètres <i>compacité, contraste et texture</i>	8
9	Valeur des prédictions en fonction des paramètres <i>compacité, contraste et texture</i>	9
10	Valeur des prédictions en fonction des paramètres <i>compacité, contraste et texture</i>	9

1 Introduction

L'objectif de ce projet est d'explorer différentes extensions de la classification SVM en utilisant des données de test. Nous étudierons différents extensions de la classification SVM, la SVM à noyau gaussien et la SVM linéaire à marge souple. En résumé, nous chercherons à comprendre ces formulations, à optimiser la SVM à noyau gaussien et à analyser la flexibilité de la SVM linéaire à marge souple en 2D et en 3D.

2 Extension-1 : SVM à noyau gaussien

La méthode utilisée dans cette implémentation de la SVM consiste à calculer une matrice de Gram pour évaluer la similarité entre les échantillons. Ensuite, des contraintes sont définies en utilisant cette matrice de Gram et les étiquettes de classe pour garantir une séparation adéquate des classes. Un problème d'optimisation quadratique est résolu pour obtenir les poids optimaux. Les vecteurs de support, qui sont les échantillons les plus pertinents pour la classification, sont sélectionnés en fonction des poids obtenus. Enfin, le biais est calculé pour déterminer la position de la frontière de décision.

On a donc implémenté la fonction SVM_3 qui permet de résoudre le problème d'optimisation permettant de chercher le maximum de la fonction \hat{L} définie par :

$$\hat{L}(\alpha_1, \dots, \alpha_n) = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i y_i K(x_i, x_j) y_j \alpha_j + \sum_{i=1}^n \alpha_i$$

avec la matrice de Gram :

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

Le résultat obtenu avec une valeur de $\sigma = 0.006$ est le suivant :

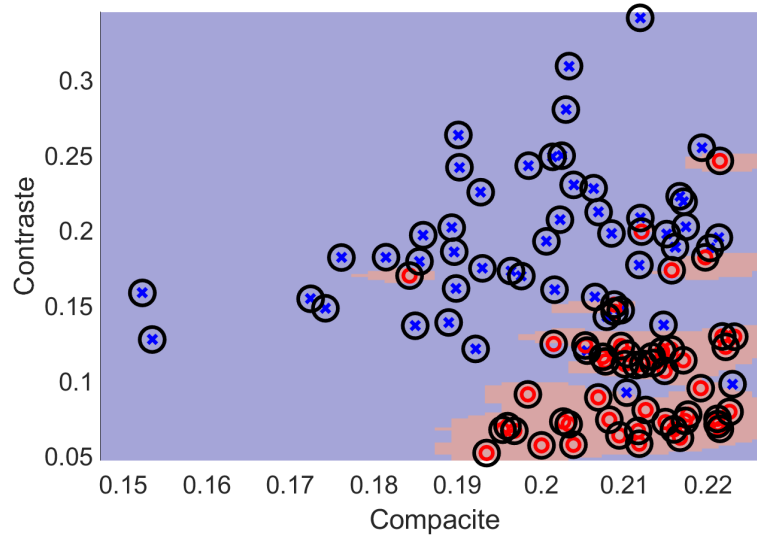


FIGURE 1 – Résultats de la classification à l'aide de la fonction SVM_3

Lors de l'utilisation de la SVM à noyau gaussien pour classer des données non linéairement séparables, nous avons observé que la classification n'était pas entièrement parfaite. Cependant, nous avons néanmoins réussi à mettre en œuvre une classification partielle qui s'avère être assez précise.

3 Extension-2 : Optimisation du SVM à noyau gaussien

Dans cette section, nous avons tracé le pourcentage de bonnes classifications en fonction de la variation des valeurs de sigma, qui ont été ajustées de 0.0001 à 0.01 avec un pas de 0.0001.

La courbe résultante nous permet de déterminer la valeur optimale de sigma qui maximise le pourcentage de bonnes classifications, ainsi que le pourcentage de bonnes classifications correspondant à cette valeur optimale.

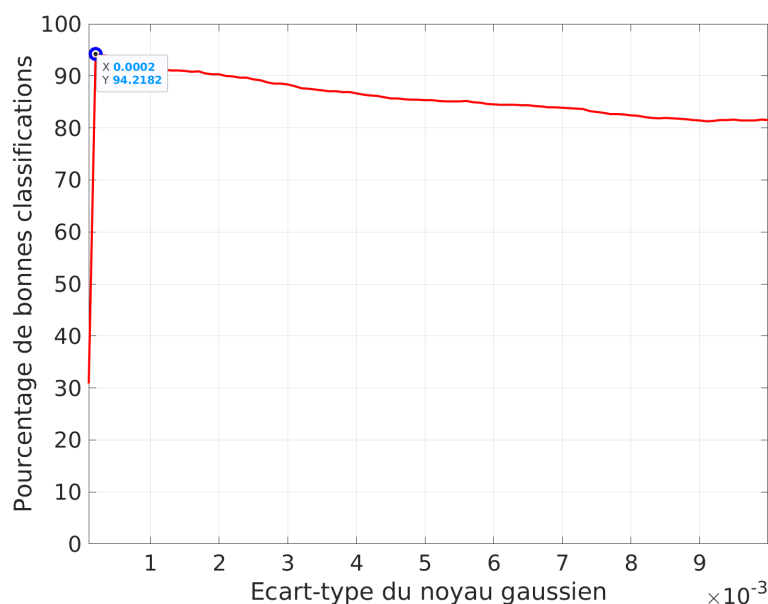


FIGURE 2 – Tracé du pourcentage de bonnes classifications en fonctions de σ

On a ainsi obtenu le résultat suivant :

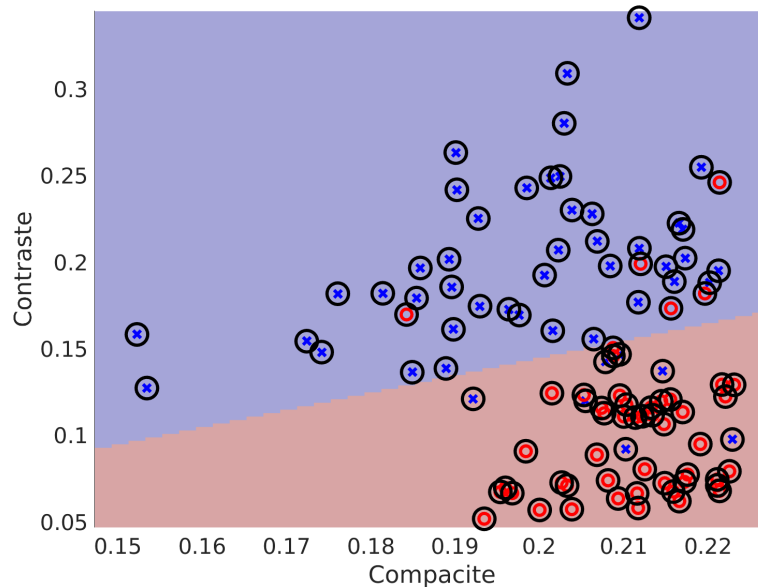
$$\begin{cases} \sigma_{max} = 0.0002 \\ Pourcentage_{max} = 94.2182 \% \end{cases}$$

Dans ce cas, il semble que le classifieur SVM à noyau gaussien avec une valeur de sigma égale à 0.0002 ait réussi à classer la plupart des échantillons avec une précision maximale de 94.2 %. Cela indique que le modèle a démontré une bonne capacité de généralisation en classant correctement la majorité des exemples lors de l'évaluation.

4 Extension-3 : SVM lineaire à marge souple

Dans cette étude, nous avons exploré l'utilisation du concept de marge souple (soft margin) pour la classification des données de manière non linéaire à l'aide d'un classifieur SVM.

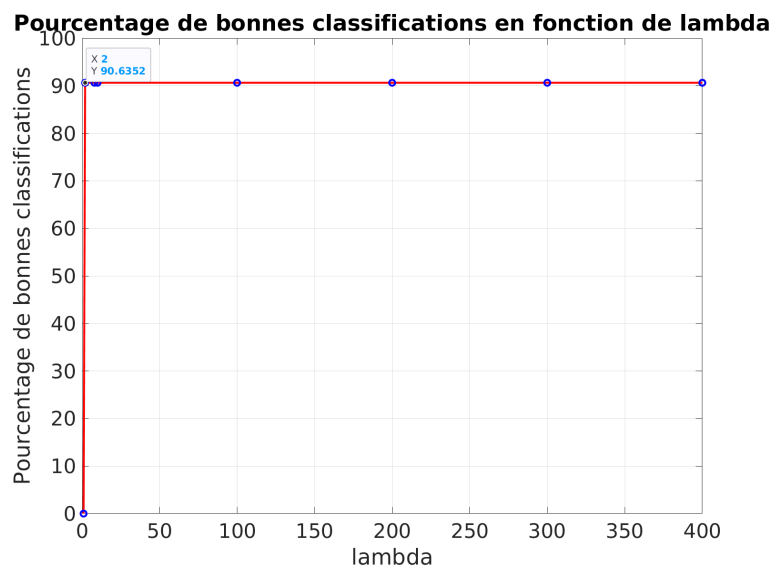
Le résultat de la classification obtenue pour une valeur de lambda $\lambda = 100$ avec un pourcentage de bonnes classifications $Pourcentage = 90.6352 \%$ est le suivant :

FIGURE 3 – Résultats de la classification SVM à marge souple λ

On a après effectué le tracé du pourcentage de bonnes classifications en fonction des valeurs de λ variant de 1 à 400 avec un pas de 1. Nous avons ainsi obtenu que :

$$\forall \lambda \geq 2, \text{Pourcentage} = \text{Pourcentage}_{max} = 90.6352 \%$$

La figure tracé envisageant le résultat obtenu montre qu'à partir du rang 2 de λ mentionné précédemment, le pourcentage de bonnes classifications est maximale.

FIGURE 4 – Tracé du pourcentage de bonnes classifications en fonctions de λ

5 Extension-4 : SVM à noyau gaussien et marge souple

Dans cette section, nous allons fusionner les deux extensions précédentes, à savoir l'utilisation du noyau gaussien dans la SVM et l'application d'une marge souple, afin de créer une nouvelle extension de classification SVM plus optimale. Le pourcentage de classification obtenu grâce à cette approche améliorée est de :

$$\text{Pourcentage}_{max} = 94.6254 \%$$

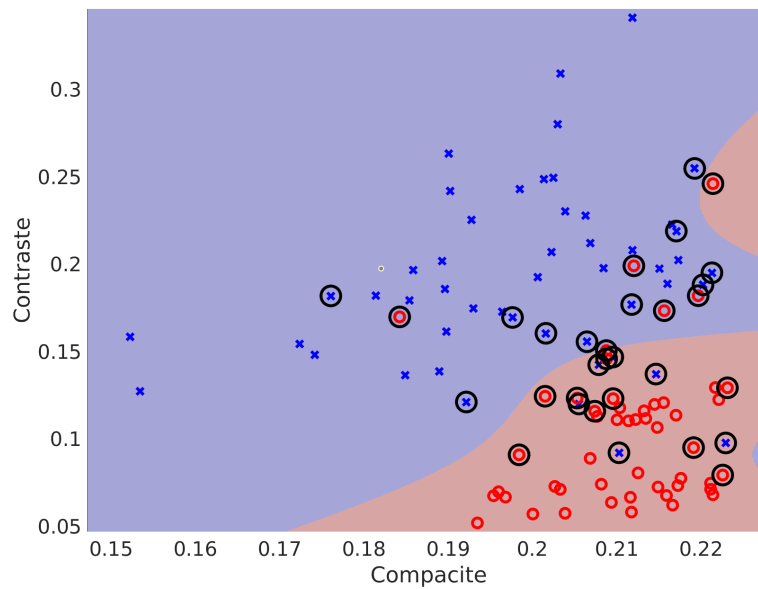


FIGURE 5 – Tracé du pourcentage de bonnes classifications en fonctions du couple λ et σ

On a effectué le tracé de la figure en 3D du pourcentage de bonnes classifications en fonction de σ et de λ , ainsi, on a obtenu la figure suivante qui modélise les points maximale trouvé :

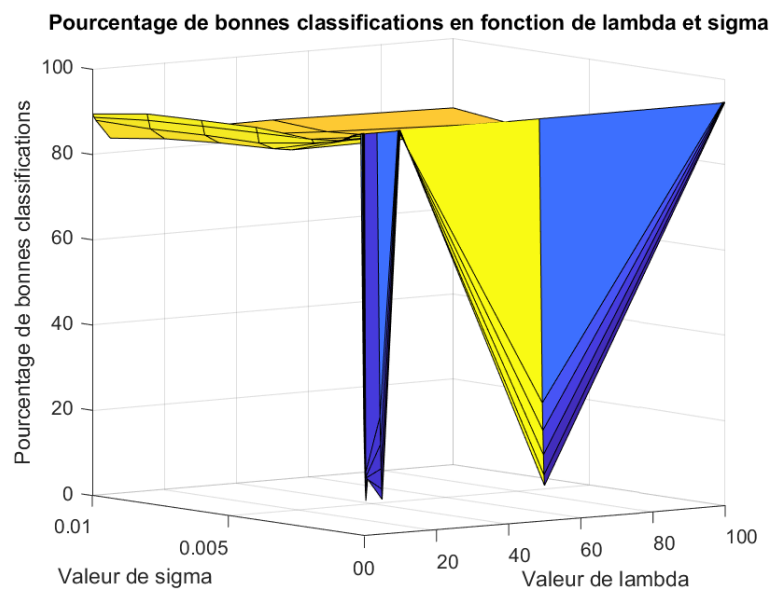


FIGURE 6 – Tracé du pourcentage de bonnes classifications en fonctions du couple λ et σ

Les couple des valeurs de σ et λ maximisant ainsi le pourcentage de bonne classifications sont :

lambda	sigma
0.10	1.0e-05
0.10	3.0e-05
0.10	5.0e-05
10	1.0e-05
10	3.0e-05
10	5.0e-05
100	1.0e-05
100	3.0e-05
100	5.0e-05

Ensuite, on a effectué le tracé de λ en fonction de σ pour en déduire une approche de la propor-

tionnalité des deux :

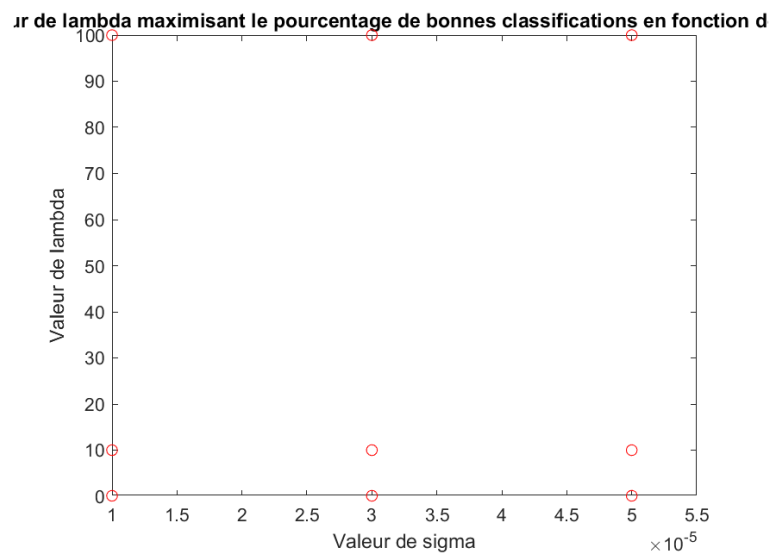


FIGURE 7 – Tracé de λ_{max} en fonction de σ_{max}

Cependant, nos observations révèlent qu'il n'y a pas de lien proportionnel entre les valeurs de σ et de λ , ce qui rend impossible d'inférer des conclusions sur la façon dont le pourcentage de bonnes classifications varie en fonction des variations du couple (σ, λ) .

6 Extension-5 : Classification de données en dimension 3

6.1 SVM à noyau gaussien en 3D

Dans cette section, nous avons abordé le problème de la classification SVM à noyau gaussien dans un scénario à trois classes, en utilisant les caractéristiques de compacité, de contraste et de texture. Les résultats obtenus montrent un pourcentage de bonnes classifications de :

$$Pourcentage_{max} = 73.9 \%$$

Les tracés montrant la valeurs des prédictions en fonction des valeur de la compacité, contraste et la texture :

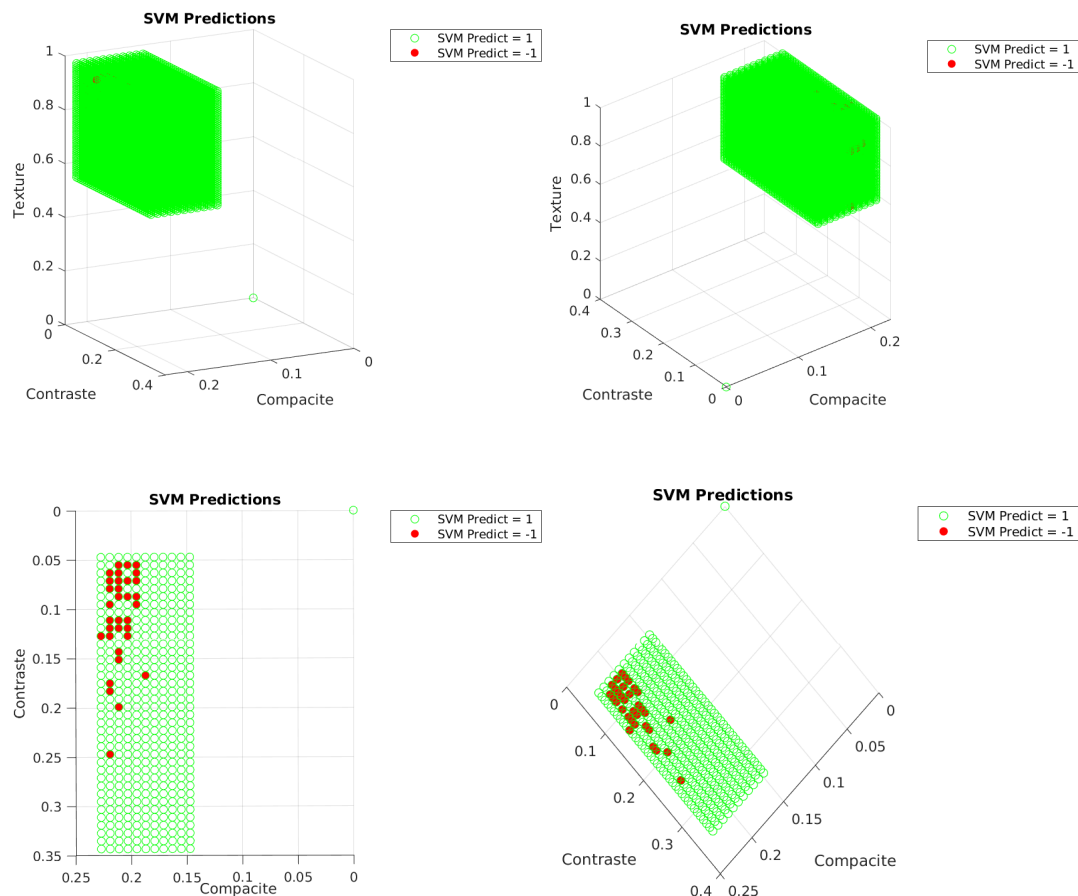


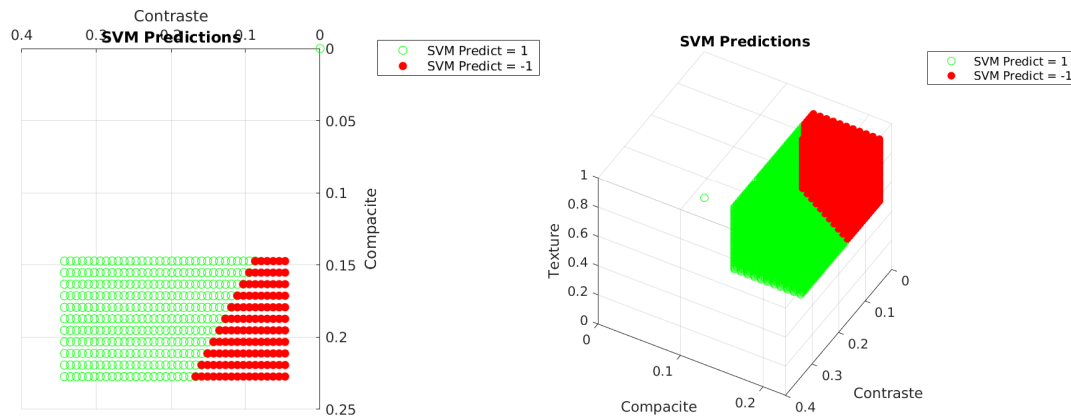
FIGURE 8 – Valeur des prédictions en fonction des paramètres *compacité*, *contraste* et *texture*

Ce pourcentage de bonnes classifications est encourageant, mais il est important de considérer d'autres solutions afin de maximiser les prédictions correctes.

6.2 SVM à marge souple en 2D

Dans cette partie du projet, nous allons mettre en œuvre la classification SVM à marge souple pour traiter une tâche de classification en trois dimensions (3D). La classification SVM à marge souple est une extension de la SVM linéaire qui permet de tolérer un certain nombre d'erreurs de classification, ce qui la rend plus flexible et adaptée à des problèmes complexes.

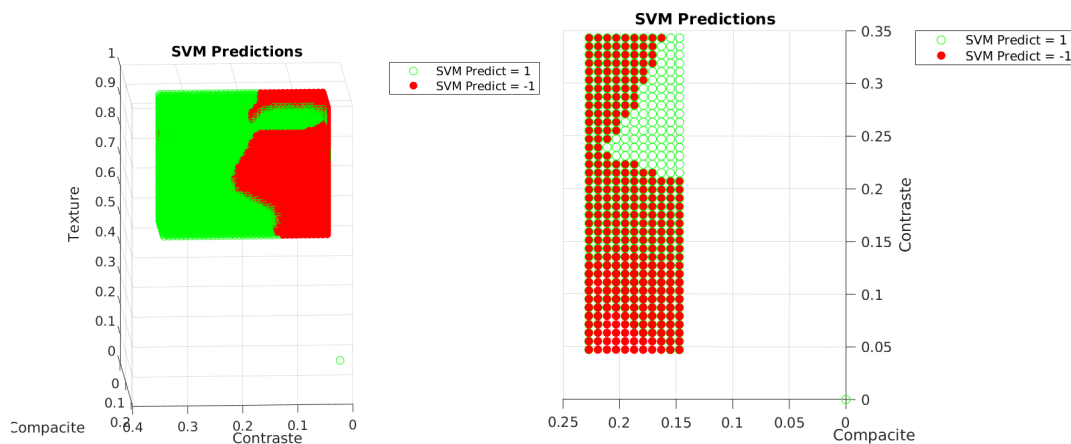
Nous avons ainsi tracé les valeurs des prédictions en fonction des paramètres *compacité*, *contraste* et *texture*.

FIGURE 9 – Valeur des prédictions en fonction des paramètres *compacité*, *contraste* et *texture*

6.3 SVM à marge souple et noyau gaussien en 3D

Dans cette section, nous avons combiné les techniques précédentes, à savoir la SVM à marge souple et la SVM à noyau gaussien, dans le but d'optimiser notre modèle de classification. En résultat, nous avons obtenu un pourcentage plus élevé de bonnes classifications, qui s'élève à :

$$\text{Pourcentage}_{max} = 83\%$$

FIGURE 10 – Valeur des prédictions en fonction des paramètres *compacité*, *contraste* et *texture*

Cette fusion des techniques a permis d'améliorer les performances de classification et de mieux prédire les étiquettes de classe pour nos données. Cela démontre l'efficacité de cette approche combinée pour obtenir des résultats plus précis et fiables.

7 Conclusion

En conclusion, ce projet visait à explorer différentes extensions de la classification SVM (Support Vector Machine) en utilisant des données de test. Nous avons étudié la formulation primale et duale de la SVM pour le séparateur linéaire, ainsi que la SVM à noyau gaussien. Nous avons optimisé la SVM à noyau gaussien en ajustant les paramètres du noyau et en sélectionnant les vecteurs de support. Enfin, nous avons examiné la SVM linéaire à marge souple. Chacune de ces extensions a permis d'améliorer la capacité de classification de la SVM.