



UNIVERSITÉ DE MONTPELLIER
Département de Mathématiques Appliquées

TP3 : Support Vector Machine

Apprentissage Statistique — HAX907X

Réalisé par :
ATTOUMANI Ibrahim



Année Universitaire 2025 – 2026

Table des matières

1. Introduction	2
2. Classification sur les donnée Iris	3
2.1. SVM linéaire : discrimination entre les classes 1 et 2 d'Iris	3
2.1. SVM polynomial : discrimination entre les classes 1 et 2 d'Iris	4
Annexe	5

1. Introduction

Les **SVM** (Support Vector Machines), introduits par Vapnik, sont des méthodes de classification très utilisées, en particulier pour la classification binaire. Elles reposent sur la recherche d'une règle de décision linéaire sous la forme d'un **hyperplan séparateur**. Pour traiter des problèmes plus complexes, cette recherche est effectuée non pas directement dans l'espace des données initiales, mais dans un **espace de caractéristiques** de grande dimension, obtenu grâce à une transformation non linéaire.

L'objectif de ce TP est d'appliquer les SVM sur des données réelles et simulées à l'aide de la librairie `scikit-learn` (qui s'appuie sur `libsvm`). Nous apprendrons également à ajuster les **hyperparamètres** et le **choix du noyau** afin de mieux contrôler la flexibilité du modèle.

2. Classification sur les donnée Iris

Dans cette section, nous allons mettre en œuvre un **SVM linéaire** afin de distinguer les classes **1 et 2** du jeu de données *iris*. Pour simplifier le problème, seules les **deux premières variables** seront utilisées. Nous conserverons la moitié des données pour l'**apprentissage** et l'autre moitié pour le **test**, afin d'évaluer la capacité du modèle à bien généraliser.

2.1. SVM linéaire : discrimination entre les classes 1 et 2 d'Iris

Ici nous allons sélectionner les classes 1 et 2 du dataset *iris*, puis conserver uniquement les deux premières variables (longueur et largeur des sépales) afin de simplifier la visualisation et la classification.

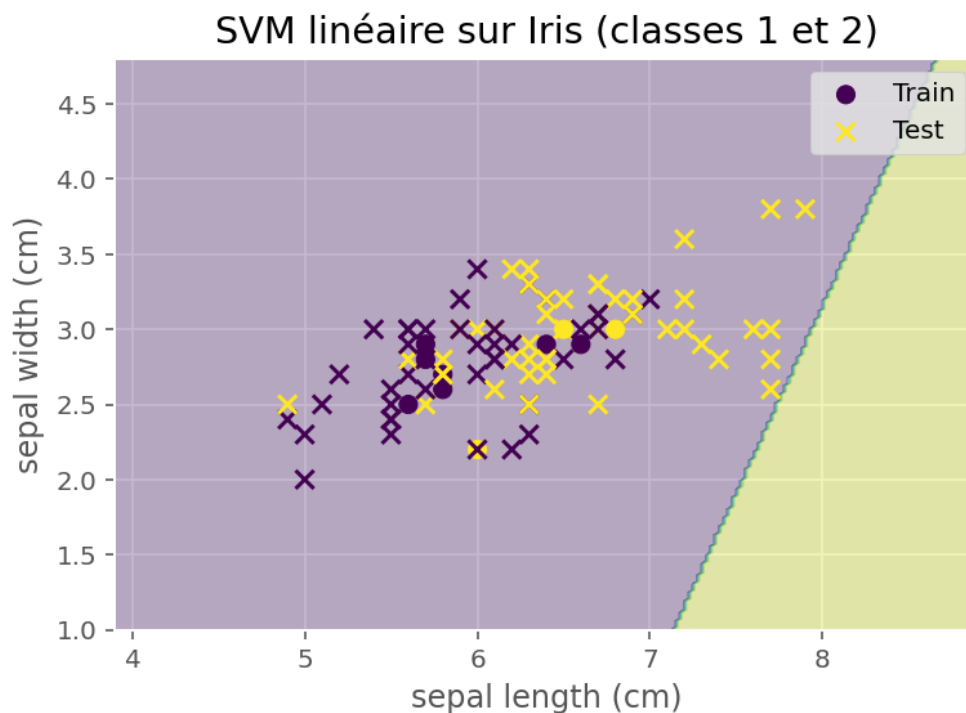


Figure 1: SVM linéaire avec 90% des données réservées au test

Ci-dessus, on a dédiées 90% des données pour le test, ce qui ne laisse qu'environ 10% pour l'entraînement (soit seulement 10 points). Avec si peu d'exemples, le SVM ne dispose pas d'assez d'informations pour bien apprendre. On a alors que la frontière de décision est très approximative et la précision sur le test chute à **0.48**, proche du hasard.

En utilisant 40% des données pour le test, nous obtenons une précision de 7%. Comparée à la **figure 1**, la frontière de décision sépare mieux les points, bien que quelques erreurs de prédiction subsistent. La **figure 2** ci-dessous illustre cette visualisation.

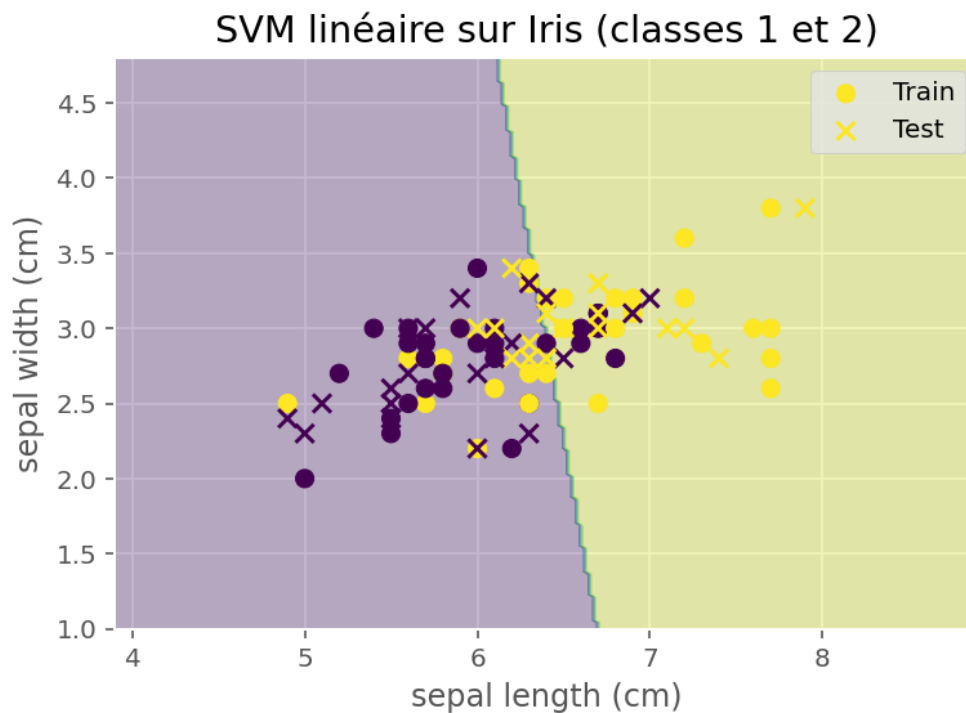


Figure 2: SVM linéaire avec 40% des données réservées au test

2.1. SVM polynomial : discrimination entre les classes 1 et 2 d'Iris

Dans cette section, nous entraînons un SVM avec noyau polynomial pour discriminer les classes 1 et 2 d'Iris. Ce noyau permet de modéliser des frontières de décision non linéaires, plus flexibles que celles du SVM linéaire. Nous comparerons ensuite ses performances et sa frontière de décision avec celles du SVM linéaire présenté précédemment.

Annexe

References

- [1] scikit-learn developers. *Support Vector Machines* — *scikit-learn documentation*. <http://scikit-learn.org/stable/modules/svm.html>. Consulté le 24 septembre 2025.
- [2] Wikipedia. *Support vector machine*. http://en.wikipedia.org/wiki/Support_vector_machine. Consulté le 24 septembre 2025.
- [3] Wikipédia. *Machine à vecteurs de support*. http://fr.wikipedia.org/wiki/Machine_%C3%A0_vecteurs_de_support. Consulté le 24 septembre 2025.