



Images/logo.png

HAX907X Apprentissage statistique

TP : Support Vector Machines (SVM)

Marine GERMAIN
Coralie ROMANI DE VINCI

Table des matières

1 Introduction

Les *Support Vector Machines* (SVM) sont une méthode de classification supervisée qui vise à séparer différentes classes en maximisant la marge entre elles. Le principe repose sur la recherche d'hyperplans séparateurs, ce qui les rend particulièrement efficaces dans des espaces de grande dimension.

Dans ce TP, nous mettons en pratique les SVM à l'aide du package `scikit-learn`. Nous appliquerons ces méthodes à la fois sur des données simulées et sur des jeux de données réels, en particulier le dataset `iris` et une base de données d'images. Nous analyserons également l'influence des noyaux et des hyperparamètres sur la performance des modèles, afin de mieux comprendre cette approche de classification.

2 Base de données `iris`

Cette section est consacrée à l'étude de la base de données `iris` sur laquelle nous ferons une étude de classification sur les classes 1 et 2. Nous considérerons d'abord le noyau linéaire puis le noyau polynomial et nous les comparerons. Le dataset est divisé en deux parties : un ensemble d'entraînement (75% des données) et un ensemble de test (25%).

2.1 Classification avec noyau linéaire (Question 1)

La classification des deux premières variables avec un noyau linéaire a pour objectif de trouver un hyperplan qui sépare au mieux les deux classes en maximisant la marge entre les points des deux classes. Les scores obtenus pour cette méthode sont les suivants :