



---

# HAX907X - Apprentissage statistique

TP : Support Vector Machines (SVM)

---

Marine GERMAIN  
Coralie ROMANI DE VINCI

# Table des matières

<b>1</b>	<b>Base de données Iris</b>	<b>2</b>
1.1	Classification avec noyau linéaire . . . . .	2
1.2	Classification avec noyau polynomial . . . . .	2
<b>2</b>	<b>SVM GUI</b>	<b>3</b>
<b>3</b>	<b>Classification des visages</b>	<b>3</b>
3.1	Influence du paramètre de régularisation . . . . .	3
3.2	Variable de nuisances . . . . .	3
3.3	Réduction de dimensions . . . . .	3

# 1 Base de données Iris

Dans cette section, nous étudierons la base de données Iris sur laquelle nous ferons une étude de classification sur les classes 1 et 2. Nous utiliserons d'abord le noyau linéaire puis le noyau polynomial et nous les comparerons. Nous avons séparé le dataset en deux parties : un ensemble d'entraînement (75% des données) et un ensemble de test (25% des données).

## 1.1 Classification avec noyau linéaire

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

**Generalization score for linear kernel: 0.7733333333333333, 0.64**

Figure 1

Pour les données d'entraînement, le modèle a classifié correctement 77,3% des données, contre 64% pour les données de test.

## 1.2 Classification avec noyau polynomial

La classification des deux premières variables avec un noyau polynomial a pour objectif de trouver une frontière de décision non linéaire capable de séparer au mieux les deux classes. Les scores obtenus pour cette méthode sont les suivants :

**Generalization score for polynomial kernel: 0.76, 0.6**

Figure 2

Pour les données d'entraînement, le modèle a classifié correctement 76% des données, contre 60% pour les données de test.

### Comparaison des deux méthodes :

- Nous disposons d'abord des scores obtenus en figure 1 et figure 2. On remarque que le score pour le noyau linéaire est plus élevé que celui du noyau polynomial, autant pour les données d'entraînements que pour les données de test.

- Ensuite, nous avons tracé graphiquement ces résultats à l'aide des frontières :  
Conclusion.

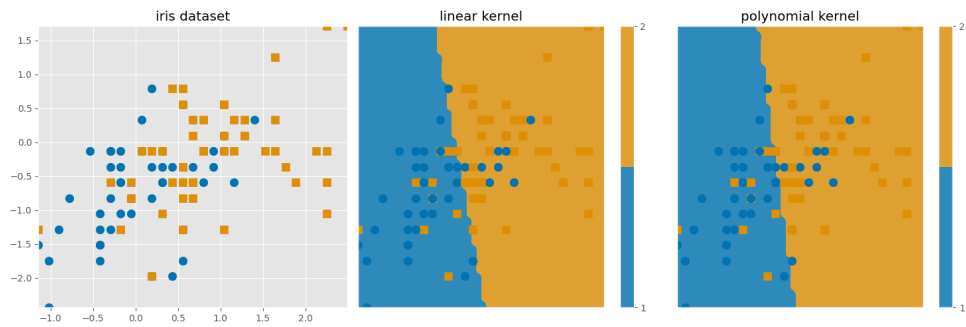


Figure 3

## 2 SVM GUI

## 3 Classification des visages

### 3.1 Influence du paramètre de régularisation

### 3.2 Variable de nuisances

### 3.3 Réduction de dimensions