



# Travaux Pratiques — Support Vector Machines (SVM)

École des Mines de Saint-Étienne

Dr. Youssef SALMAN

31 octobre 2025

## Objectif du TP

L'objectif de ce TP est de comprendre et d'illustrer expérimentalement les notions suivantes :

- marge dure et marge souple (paramètre  $C$ ) ;
- vecteurs de support et largeur de marge ;
- astuce du noyau (*kernel trick*) et classification non linéaire.

Les exercices sont réalisés en **Python** avec les bibliothèques NumPy, Matplotlib et **scikit-learn**.

---

## Question 1 — SVM linéaire : marge dure et marge souple

- 1) Générez un jeu de données 2D quasi linéairement séparable (`make_classification`).
- 2) Entraînez trois SVM linéaires avec  $C \in \{0.1, 1, 1000\}$ .
- 3) Tracez la frontière de décision, les marges  $\pm 1$  et les vecteurs de support.
- 4) Calculez et commentez la largeur de la marge  $\frac{2}{\|\beta\|}$ .

**Question de réflexion :** que se passe-t-il lorsque  $C$  est petit ou très grand ?

---

## Question 2 — Validation croisée et performance

- 1) Utilisez `make_moons` pour créer des données non linéaires bruitées.
- 2) Séparez le jeu en apprentissage et test (70%/30%).
- 3) Réalisez une validation croisée (5-fold) sur  $C \in \{0.01, 0.1, 1, 10, 100\}$  pour un SVM linéaire.
- 4) Évaluez le modèle optimal sur l'ensemble test : **accuracy**, matrice de confusion, rapport de classification.

**Question de réflexion** : pourquoi le modèle linéaire est-il limité pour des données non linéaires ?

---

## Question 3 — Astuce du noyau : polynôme et RBF

- 1) Générez un jeu de données `make_circles` (données non linéaires).
- 2) Comparez par validation croisée :
  - un SVM **polynomial** (degré  $p = 2, 3, 4$ ),
  - un SVM **RBF** (paramètres  $C$  et  $\gamma$ ).
- 3) Tracez les frontières de décision du meilleur modèle.
- 4) Comparez le nombre de vecteurs de support et la performance.

**Question de réflexion** : expliquez le rôle du noyau et de ses paramètres (**degré** pour le polynôme,  $\gamma$  pour le RBF).

---

## Méthodologie et livrables

- Documentez chaque étape : génération de données, apprentissage, affichage graphique, interprétation.
  - Comparez vos résultats avec les notions théoriques du cours : **marge**, **vecteurs de support**, **influence de  $C$** , **noyaux**.
  - Rendez :
    - le code Python commenté ;
    - les graphiques obtenus ;
    - une courte analyse par question.
-

## Rappel méthodologique du SVM

- **Marge** : distance entre les deux hyperplans  $h(x) = \pm 1$ , donnée par  $\frac{2}{\|\beta\|}$ .
- **Paramètre  $C$**  : équilibre entre largeur de marge et erreurs de classification.
- **Noyaux usuels** :

Linéaire :  $K(x_i, x_j) = \langle x_i, x_j \rangle$

Polynôme :  $K(x_i, x_j) = (c + \langle x_i, x_j \rangle)^p$

Gaussien :  $K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$

*Fin du TP — Bon travail et bonne exploration !*