

TP: SUPPORT VECTOR MACHINE

1. Prétraitement des Données

- a) Charger le jeu de données choisi (choisir une base de données) via scikit-learn.
- b) Si le jeu de données contient plus de 2 caractéristiques:
 - Soit réaliser une sélection des 2 caractéristiques les plus discriminantes (par exemple à l'aide d'analyses exploratoires),
 - Soit utiliser une méthode de réduction de dimensionnalité comme le PCA **uniquement pour la visualisation** (le modèle SVM doit être entraîné sur les données d'origine).

2. Séparation du Jeu de Données

- a) Diviser le jeu de données en ensemble d'entraînement (70 %) et ensemble de test (30 %).
- b) Assurer la reproductibilité en utilisant un `random_state` fixé.

3. Modélisation avec SVM

- a) Entraîner un SVM non linéaire (noyau RBF) pour la classification multi-classes.
- b) Expliquer la différence avec un SVM linéaire (par exemple en modifiant le paramètre kernel).

4. Optimisation des Hyperparamètres

- a) Mettre en place une recherche d'hyperparamètres avec `GridSearchCV` afin d'optimiser les paramètres `C` et `gamma`.
- b) Comparer la performance du modèle avec et sans optimisation des hyperparamètres.
- c) Interpréter les résultats (quels paramètres ont été retenus et pourquoi).

5. Analyse des Vecteurs Supports

- a) Afficher le nombre et la répartition des vecteurs supports par classe.
- b) Expliquer pourquoi certains points sont considérés comme vecteurs supports et leur impact sur la frontière de décision.

6. Visualisation des Frontières de Décision

- a) Utiliser une bibliothèque de visualisation pour tracer les frontières de décision sur l'ensemble d'entraînement.
- b) Ajouter un tracé spécifique des vecteurs supports sur le graphique.
- c) Si le modèle est entraîné sur plus de 2 dimensions, utiliser une projection (par exemple PCA uniquement pour la visualisation) pour représenter la frontière.

7. Évaluation du Modèle

- a) Calculer les métriques d'évaluation (accuracy, précision, rappel, F1-score) sur l'ensemble de test.
- b) Comparer les performances obtenues avec différentes configurations (par exemple SVM linéaire vs non linéaire, avec ou sans optimisation).