

**Résumé:**

La présence des données mal classées dans les problèmes de classification affecte généralement la performance du modèle. Nous proposons une nouvelle approche visant à identifier et supprimer les données incorrectement classées de l'ensemble d'apprentissage, afin d'améliorer l'exactitude des modèles de prédiction sur les ensembles tests. Notre approche utilise l'algorithme d'apprentissage SVM avec une perte quadratique pour créer des classificateurs qui permettent de filtrer des données mal classées de l'ensemble d'apprentissage. Nous utilisons le classificateur filtre qui fournit le meilleur pourcentage de classification sur intervalle  $\mathcal{C} \in [\mathcal{C}_{min}; \mathcal{C}_{max}]$ . Les performances du classificateur SVM optimal ont été comparées à celles des modèles SVM utilisant différentes valeurs du paramètre  $\mathcal{C}$  sur trois ensembles de données de l'UCI. Pour cette évaluation, nous utilisons le taux de bonne classification, qui est la mesure de performance la plus courante dans les problèmes de classification. Les résultats obtenus montrent que le classificateur optimal en termes de pourcentage de classification est un filtre efficace pour la détection des données incorrectement classées.

**Mots-clés:** optimisation convexe; algorithme de Frank -Wolfe; classification; SVM.

**Abstract:**

The presence of misclassified data in classification problems generally degrades model performance. We propose a novel approach for identifying and removing misclassified samples from the training set in order to improve the predictive accuracy of classification models on test data. Our approach employs a Support Vector Machine (SVM) with the squared loss to build classifiers capable of filtering misclassified training samples. The filtering classifier is selected as the one achieving the highest classification accuracy over the interval  $\mathcal{C} \in [\mathcal{C}_{min}; \mathcal{C}_{max}]$ . The performance of the optimal SVM classifier was compared with that of SVM models using different values of the parameter  $\mathcal{C}$  on three binary classification datasets from the UCI Machine Learning Repository. Classification accuracy, the most widely used performance metric for classification tasks, was adopted as the evaluation criterion. The experimental results demonstrate that the optimal classifier, in terms of classification accuracy, provides an effective filtering mechanism for detecting misclassified training samples.

**Keywords:** convex optimization; Frank -Wolfe algorithm; classification; SVM.