

Probabilistic Approach to One-Class Support Vector Machine

Vincent Leclère

CERMICS, École des Ponts Paristech, France

Laurent El Ghaoui

UC Berkeley, California

Edouard Graves

UC Berkeley, California

Mots-clefs : Machine Learning, Classification, Support Vector Machine

Classification is one of the main problem addressed by machine learning algorithms. Among them the Support Vector Machine has attracted a lot of interest and shown success in the past decades. Support Vector Machines (SVM) are originally tailored for binary classification. However sometimes we have few example of negative dataset, and we can turn to one-class SVM. In this presentation we propose a probabilistic interpretation of the one class Support Vector Machine approach, and an extension especially adapted in the case of highly imbalanced dataset. Indeed, we consider a binary classification problem where we represent the negative dataset by its two first moments, while still modeling the positive class by individual examples. The optimization problem is shown to have an equivalent formulation to a one class SVM applied to the positive dataset after some preprocessing. The usual one class SVM corresponding to the case where the negative class has mean 0 and identity variance.

We demonstrate empirically, on a protein classification task and a text classification task, that our approach achieves similar statistical performance than the two mainstream approaches to imbalanced classification problems, while being more computationally efficient.