One class learning as a null space kernel Rayleigh quotient. Application to abnormal events detection in video sequences

Dufrenois Franck[1]

[1] LISIC- 50 rue Ferdinand Buisson- 62100 Calais France

Abstract

Many problems in machine learning can be simplified as the maximization (minimization) of a generalized Rayleigh quotient, given by:

$$\max_{\boldsymbol{w}} \rho(\boldsymbol{w}; Q, P) = \frac{\boldsymbol{w}^T Q \boldsymbol{w}}{\boldsymbol{w}^T P \boldsymbol{w}} \quad (1) \tag{1}$$

where Q and P are positive definite matrices, and $\boldsymbol{w} \neq 0$ is an optimal projection into a lower-dimensional space which solves the problem (1). Two classical applications formulated as (1) are principal composant analyis (PCA) and linear discriminant analysis (LDA). In order to deal with non normal data distributions, a kernel embedding is commonly performed which generalizes the previous criterion for any kind of distributions. Recently, we proposed a new kernel Rayleigh quotient for solving the one class learning problem [1]. Unlike binary/multi-class classification methods, one class classification tries to isolate a target (positive) class when the negative class is either poorly represented, even not at all or is not well statistically defined. Our formulation introduces two regularized specific scatter matrices $Q(\boldsymbol{y})$ and $P(\boldsymbol{y})$ which are parameterized by an unknow binary vector $\boldsymbol{y}$ determining the membership of a data to the positive or negative class. We show in [1] that the optimal separation of these two data populations amounts then to achieve two joint actions: *dimensionality reduction* and *classification*. However, several aspects limit its use: firstly, the accuracy of the dimensionality reduction depends on the representativeness of the abnormal data which can be low for specific data sets. Secondly, when the data dimensionality becomes much greater than the training sample size, the method can be badly conditionned (singularity problem).

In this paper, we introduce a null space based extension of this criterion. The principle of this extension is to introduce a joint subspace where the training target data set has zero covariance. Then, a simple distance measure can be derived in this subspace to decide about abnormality of a test data. We show here that this formulation implicitly avoids the singularity problem for under sampled data sets. We show also that the dimension of this specific subspace is directly linked to the contamination rate which is of course unknown. This issue is solved by introducing a single artificial counter-example. This operating strategy allows both to reduce the size of the null space to a single null direction which can be estimated from a simple power method and to maintain an enough generalization performance. A comparative study with different one class learning methods is conducted both on moderated and high dimensional data sets. An original application to the detection of abnormal events in video sequences is also presented.

References

[1] F. Dufrenois and J.C. Noyer. One class proximal support vector machines. Pattern Recognition. 52, pp. 96–112, 2016.