

One-Hit Multiple Instance Logistic Regression for Binary Classification and Its Application to Atomic Force Microscopy Images for Bladder Cancer Determination

Authors : Eugene Demidenko, John Seigne, Igor Sokolov

Abstract : Multiple instance classification is a known machine learning technique when only a bag of features is labeled. The method of binary multiple instance classification, termed multiple instance logistic regression (LR), received the most attention as a well-defined statistical model. This algorithm is realized in several computer languages, including R (milr) and MATLAB. This work suggests improving this model, which is called the one-hit multiple instance LR. Unlike the existing approach, where unknown labels are treated as missing observations, our model directly implements the ML approach. As such, it is methodologically straightforward and computationally stable, especially when features are highly correlated and/or bags are heterogeneous. Since the one-hit LR admits a closed form for the log-likelihood function, an efficient Fisher scoring algorithm applies with the variances of the regression coefficients computed through the inverse of the Fisher information matrix at the final iteration. Numerical experiments demonstrate the superiority of the one-hit LR in terms of regression coefficients and classification accuracy. Another advantage of our approach is developing the optimal probability threshold for classification (the traditional threshold equals 0.5). The one-hit LR is illustrated with a noninvasive bladder cancer identification where each patient, in the multiple instance terminology 'bag,' contains feature images of multiple cells from a urine sample of the same individual. We show that the one-hit LR with two Atomic Force Microscopy (AFM) image features leads to a perfect (AUC=1) or almost perfect (AUC=0.978) classification of normal and cancer patients among 20 individuals. The p-value 0.0018 confirms that the latter AUC is unlikely to be obtained by chance.

Keywords : AUC, classification accuracy, classification p-value, Fisher information, ML, ROC curve

Conference Title : ICAMME 2025 : International Conference on Applied Mathematics and Mathematical Engineering

Conference Location : London, United Kingdom

Conference Dates : June 26-27, 2025