عنوان مقاله:

Outlier Detection for Support Vector Machine using Minimum Covariance Determinant Estimator

نویسندگان:

.M. Mohammadi - *Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran*

.M. Sarmad - *Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran*

خلاصه مقاله:

The purpose of this paper is to identify the effective points on the performance of one of the important algorithm of data mining namely support vector machine. The final classification decision has been made based on the small portion of data called support vectors. So, existence of the atypical observations in the aforementioned points, will result in deviation from the correct decision. Thus, the idea of Debruyne's outlier map is employed in this paper to identify the outlying points in the SVM classification problem. However, due to the computational reasons such as convenience and rapidity, a robust Mahalanobis distance based on the minimum covariance determinant estimator is utilized. This method has a good compatibility by the data with low dimensional structure. In addition to the classification accuracy, the margin width is used as the criterion for the performance assessment. The larger margin is more desired, due to the higher generalization ability. It should be noted that, by omission of the detected outliers using the suggested outlier map the generalization ability and accuracy of SVM are increased. This leads to the conclusion that the proposed method is very efficient in identifying the outliers. The capability of recognizing the outlying and misclassified observations for this new version of outlier map has been retained similar to the older version, which is tested on the simulated and real world data.

کلمات کلیدی:

Support Vector Machine, Outlying/Misclassified points, Robust statistics, Mahalanobis Distance, Minimum Covariance Determinant estimator