SVM sensitivity compared to others

Michael Belias

12/20/2021

Abstract

Background: Noise in data is a topic of concern in Machine learning field.

Objective: To assess the sensitivity of SVM compared to popular machine learning algorithms under different scenarios with feature/attribute noise.

Methods: We simulated data with different scenarios Results:

Conclusion:

Introduction

Real-world data often may contain noise from various sources. However, machine learning methods are developed under the assumption that data are clean. Applying them on noisy data may produce inaccurate models. However, not all methods are expected to be equally affected. For instance,

This paper is organised as follows. In the methods section we shortly introduce machine learning methods and noise definitions. We provide the simulation algorithm and setup. In the results section we show the performance of the methods mentioned above. Finally, in the conclusion section we provide conclusions about the performance of the algorithms on different data-sets and various levels of noise.

Methods

Real-world data may contain noise. However, noise is often related to the way data were gathered and processed. In that sense noise may come from different sources and may appear in different forms. Nettleton et al. [1] advocates that the two main sources are measurement error and random error. Measurement error is the difference between a measured quantity and its true value , while random error is noise occurring due to the random nature of an experiment. Another type of noise may be due to missing data.

To assess the effect of feature noise we simulated data

Results

Conclusion

References

1.   Nettleton DF, Orriols-Puig A, Fornells A (2010) A study of the effect of different types of noise on the precision of supervised learning techniques. Artificial intelligence review 33:275–306