



Variable Selection for Clustering and Classification

Jeffrey L. Andrews, Paul D. McNicholas

(Submitted on 21 Mar 2013)

As data sets continue to grow in size and complexity, effective and efficient techniques are needed to target important features in the variable space. Many of the variable selection techniques that are commonly used alongside clustering algorithms are based upon determining the best variable subspace according to model fitting in a stepwise manner. These techniques are often computationally intensive and can require extended periods of time to run; in fact, some are prohibitively computationally expensive for high-dimensional data. In this paper, a novel variable selection technique is introduced for use in clustering and classification analyses that is both intuitive and computationally efficient. We focus largely on applications in mixture model-based learning, but the technique could be adapted for use with various other clustering/classification methods. Our approach is illustrated on both simulated and real data, highlighted by contrasting its performance with that of other comparable variable selection techniques on the real data sets.

Subjects: **Computation (stat.CO)**

Download:

- [PDF](#)
- [Other formats](#)

Current browse context:

stat.CO

[< prev](#) | [next >](#)

[new](#) | [recent](#) | [1303](#)

Change to browse by:

[stat](#)

References & Citations

- [NASA ADS](#)

Bookmark([what is this?](#))

