Abstract

This research investigates the effectiveness of a non-convex clustering criterion with the ability to discriminate clusters by means of quadratic boundaries that take into account cluster variances. Since no algorithms have been shown to work efficiently and effectively for this kind of criterion, we introduce and evaluate a generalized version of the incremental one-by-one clustering algorithm of MacQueen (1967) that is suitable for general variance-based criteria, whether convex or not. An experimental evaluation shows that the criterion performs remarkably well with a variety of heterogeneous data sets, both synthetic and real-world. Two novel applications of unsupervised learning to problems in the financial domain are then developed to test the method further. First, given a portfolio of investments with potentially hundreds of assets, we pose the problem of reducing the number of such assets while preserving its risk-return characteristics. Next, we tackle the problem of finding risk-reward opportunities to short-sell securities. For both problems, we offer novel clustering-based solutions and proceed to show that the clustering criterion addressed in the present research is an excellent choice.

Defense Committee: Casimir Kulikowski (chair), Ahmed Elgammal and Apostolos Gerasoulis