

K-Means Clustering

Preston Engstrom

Abstract—The K-Means clustering algorithm has been a mainstay of unsupervised machine learning for decades. The algorithm is explored in this report using known datasets of 2D points, using several cluster counts and distance measures, with results compared using the Dunn Index for cluster analysis.

Index Terms—Clustering, K-means, Machine Learning

I. INTRODUCTION

THE K-Means Clustering Algorithm has continued to be the workhorse of clustering tasks since its development in 1967 by James MacQueen. Clustering is the task of grouping a set of unlabeled data vectors into groups, or clusters. Items in a cluster should be more alike items in the same cluster, and much less like items in other clusters. Clustering is one of the most common tasks of exploratory data mining for extracting preliminary observations about datasets. In this implementation, the data consists of points on a 2D plain. The clusters are simply the collection of points closest to the center of the cluster based on some distance measure.

II. THE K-MEANS ALGORITHM

III. RESULTS AND DISCUSSION

IV. CONCLUSION

The conclusion goes here.