**K-Means Clustering EViews add-in**

Erhard Menker

June 6, 2017

**Overview**

This document explains the theory & motivation behind k-means clustering, and presents the application of the algorithm using the EViews add-in. The algorithm’s implementation adheres to that given by Dr. Andrew Ng’s machine learning course.[[1]](#footnote-1) However, any deviations from the pseudocode are of course the error of the add-in’s author. Inquiries of any kind are warmly welcomed; the add-in’s author, Erhard Menker, can be reached via email ([ejmenker@gmail.com](mailto:ejmenker@gmail.com)), and the working version of the project will continue to be developed on Github.[[2]](#footnote-2)

**Cluster Analysis Motivation**

Cluster analysis is an unsupervised machine learning algorithm, meaning it is applied to “unlabeled data” (classification/categorization is not included in the dataset of the observations). Given a dataset of n observations, k-means clustering assigns each observation in the dataset to belong to one of cluster centroid 1 or 2 or … or k, conditioned that the # of centroids is less than n. This assignment is done by finding the centroid that minimizes the cost function, a measure that values centroids that have close proximity to their associated points. Therefore, clustering algorithms are a useful form of exploratory data analysis and can be used in EViews with time series (e.g. classifying macroeconomic regimes over a country’s history) & cross section (e.g. customer segmentation) data.

1. <https://www.coursera.org/learn/machine-learning/lecture/93VPG/k-means-algorithm> [↑](#footnote-ref-1)
2. <https://github.com/ErhardMenker/kMeans4EViews> [↑](#footnote-ref-2)