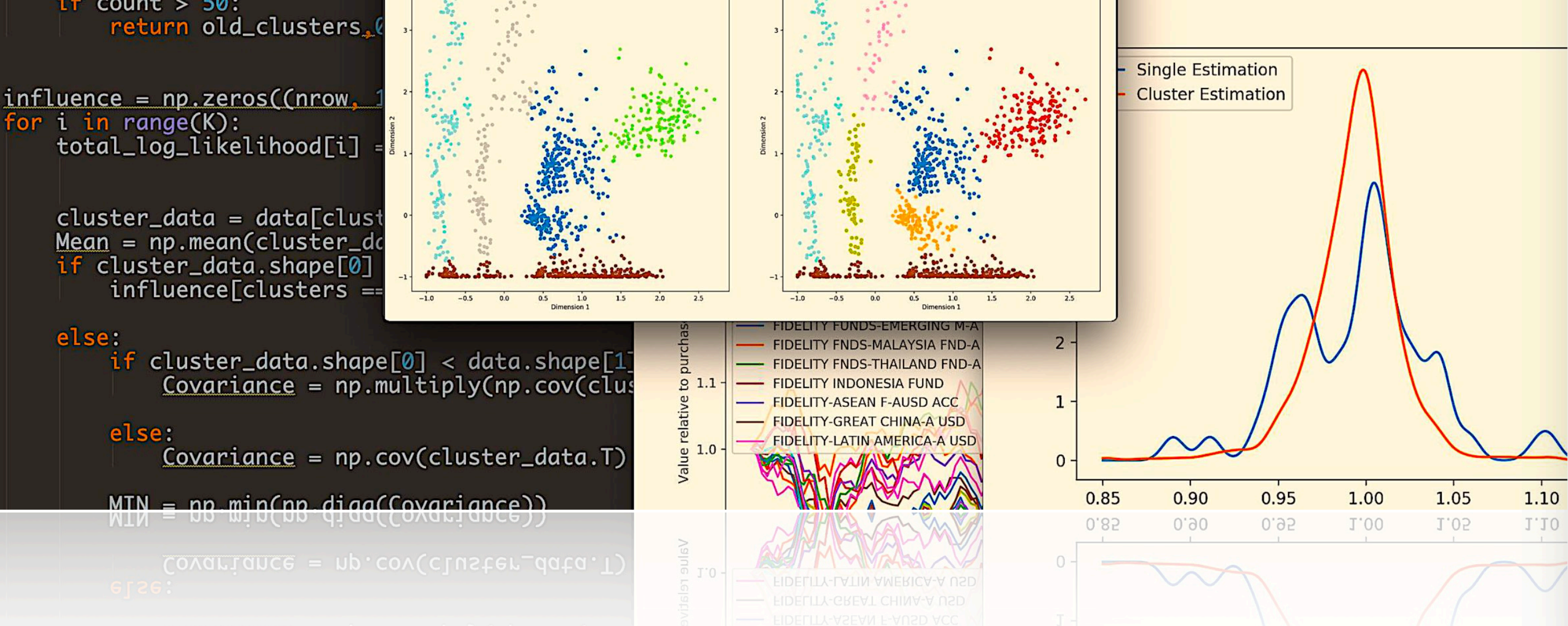

Clustering Time Series with Artificial Intelligence.

Author: *Per Wilhelmsson*

Executive summary of the Master's Thesis:
Hierarchical Clustering of Time Series using Gaussian Mixture Models and Variational Autoencoders.

Machine learning is rapidly changing the financial industry. This thesis contributes by clustering financial assets using artificial neural networks. The goal is to enhance analysis and predict the future.

Clustering is the art of finding patterns in data. The goal of clustering is to 1) put similar objects in the same group and 2) put dissimilar objects in different groups. This notion is inspired by how we divide and categorise the world in order to make sense of it. For example, we can categorise bicycles and cars into different means of transportation. In other situations, we don't distinguish between them and categorise them both as vehicles. A clustering algorithm mimics this process and tries to find meaningful categorisations for specific applications.

Clustering is widely applicable. For example, if you want to invest your money, cluster analysis can give you an edge on financial market. The market is inherently unpredictable and you always have to ask whether a price movement reflects the economical state of a company, or if the price is moving due to some random event. By looking at similar assets, you may detect anomalies and decide whether an asset is behaving reasonably or not. Furthermore, the financial market is huge and you cannot continuously analyse every individual asset. By clustering similar assets, the size of the problem shrinks and becomes more manageable.

But how do you know which assets are similar? And how do you know how many clusters there are in the market?

To answer these questions, I built an AI algorithm comprised of two machine learning models, a **Variational Autoencoder** (VAE) and a **Gaussian Mixture Model** (GMM).

The VAE is the eyes of the AI. It can see high dimensional objects and compress them into a comprehensible, low dimensional space. We can also tune how sharp it sees. For cluster analysis, we make it see blurry enough to mistake similar objects to be the same, but clear enough to distinguish more dissimilar objects. This is precisely what we want, since cluster analysis is used to extract the big picture in data.

If the VAE is the eyes of the AI, then the GMM is the brain. Just like the human brain, the GMM struggles to comprehend high dimensional objects, but with help from the VAE, it can understand them. The GMM decides what objects to cluster and how many clusters there are by fitting multiple Gaussian distributions to the data and utilising the likelihood framework. In short, the VAE transforms the objects into dots on a paper and the GMM colors them. Figure 1 shows an example.

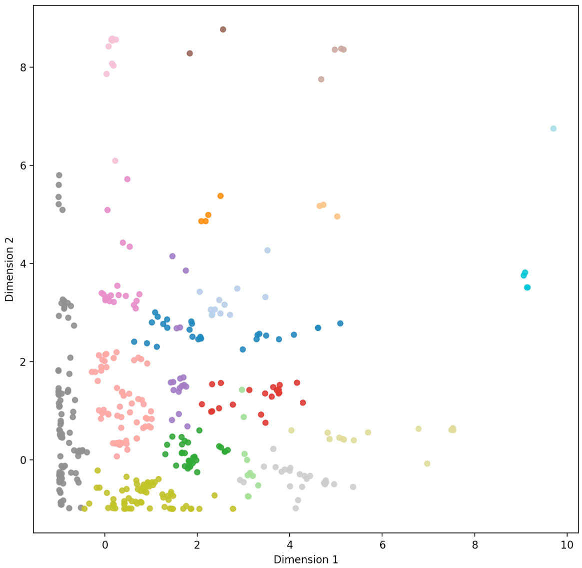

Figure 1: This picture shows 438 financial assets compressed from 100 dimensions into a 2-dimensional space. Each point is a financial asset.

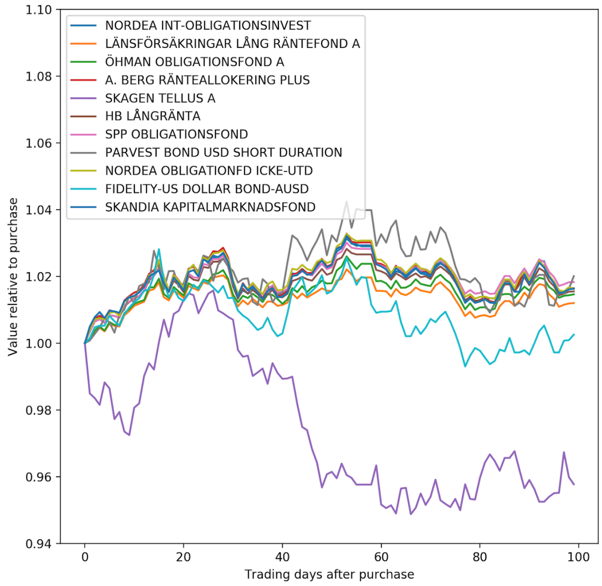

Figure 2: A cluster of 11 financial funds is shown, where the middle 8 funds are identified as stable cluster members and the other 3 funds are identified as unstable members.

Compared to similar clustering algorithms, the VAE has better vision, and the GMM is smarter. To illustrate this, the algorithm is tested on the fund market. Here, it manages to cluster the market and give information on whether to further merge or divide clusters. It also tells us which funds that are most representative of the cluster and which that are outliers. Standard clustering algorithms don't do any of this by themselves. Figure 2 shows a cluster of funds made by the algorithm. In addition to grouping the funds, the algorithm correctly identifies three outliers.

The clusters are also tested on a few predictive tasks with good results.The clusters are used to improve predictions on future price movements of individual funds and to predict which funds that are likely to change cluster assignments over time. They are also used to build robust and profitable portfolios.

In addition to being useful to investors, the algorithm is engineered to be applied to clustering of general time series. This generality and the good results above makes the algorithm a part of the paradigm shift towards the automated future.