Birds of a feather flock together? Clustering approach to identify robust economic preference types

Yu Wang*

December 24, 2024

Abstract

This study attempts to use clustering analysis to identify robust economic preference types of people with a comprehensive task-level survey dataset collected from Global Preference Survey. Specifically, this study uses generative model shallow clustering methods and deep clustering method to identify robust economic preference types. The results suggest that there are about 12 diverse and robust preference types in the dataset, with the top three types accounting for about 55 percent of the subjects.

Keywords: economic preference, risk aversion, ambiguity aversion, time preference, social preference, clustering analysis, mixture model, machine learning.

JEL Codes: C45, C52, C91, D90.

*dr.yuwang@outlook.com.

1 Introduction

Already in Ancient Greece, philosophers explored to extract and summarize the subtle individual differences in taste, personality, and behavior to understand humans. In modern economics and decision science, agents are characterized by their economic preferences over different aspects of consumption. "Preference" is the order an agent places on the alternatives according to her utility function, leading to an optimal choice in the decision-making process.

There are various facets of preferences, among which the most widely studied by the scholars are the preferences regarding risky, ambiguous, inter-temporal, or interpersonal decisions. Risk preference represents the degree people prefer risky outcomes with the knowledge of the precise probability of potential outcomes of events; ambiguity preference represents the degree people prefer measurable and known risks over unmeasurable and unknown uncertainty; time preference measures one individual's imperfect willpower or primitive impulsiveness to choose early rewards over delayed consumption; social preference (a.k.a. other-regarding preference), by relaxing the assumption of self-interest, describes one individual's intrinsic tendency and motivation to not only care about her material payoff, but also care about others' welfare and intention in social interactions.

"Birds of a feather flock together." Although the preference heterogeneity problem has been extensively investigated in theoretical and empirical research, the degree of preference homogeneity among a population has been under-explored in the literature. Some studies have attempted to apply clustering analysis to identify and characterize homogeneous subgroups among the heterogeneous population. Recent studies include using clustering analysis in studying personality trait types with representative surveys (Gerlach et al., 2018) and preference homogeneity of rural households (Chowdhury, Sutter, and Zimmermann, 2022). However, the literature had paid little attention to the preference homogeneity and heterogeneity of individuals.

This study attempts to use clustering analysis to identify robust economic preference types (i.e., clusters) of people with a comprehensive task-level dataset collected from Global Preference Survey. Clustering is a fundamental and classical machine learning task which has been extensively applied in the industry. The primary objective of clustering is to assign the data instances into groups according to the rule that similar instances are assigned to the same group and dissimilar instances are separated into different groups. Classical clustering algorithms can be classified into (1) partitive methods that assign the data instances to the closest cluster centroid in terms of distance or dissimilarity metrics, (2) hierarchical methods that assign the data instances into clusters according to the hierarchical linkages in a bottom-up or top-down approach, (3) density-based methods that assign the data instances in the area with high density into the same cluster, and (4) generative methods that assume those data instances belonging to the same cluster are generated from the same distribution, e.g., Gaussians (Zhou et al., 2022). For preference heterogeneity problems, clustering analysis can endogenously identify the various preference types in the population without any presumptions regarding the existence or the properties of preference portraits. The observed data themselves can determine what kinds of preference types exist and how the preferences for risky, ambiguous, intertemporal and interpersonal decisions are combined and intertwined (Bruhin, Fehr, and Schunk, 2019).

Specifically, this study will employ generative clustering methods to identify robust economic preference types. Generative methods can represent, model, and generate data instances, thus it draws growing attention from machine learning scholars and practitioners. A generative model assumes the cluster structures and outputs soft assignments of data instances by estimating the probability density of the observed data. There are three advantages of generative models that make them suitable for this economic preference clustering task. First, generative models make distributional assumptions of the underlying data generating process of the observed data therefore they are more interpretable and tractable. Second, generative models can return the likelihood of one instance under the model. This can help us determine if one subject is an outlier in terms of economic preferences according to his likelihood under the model. Finally, one advantage of generative model-based clustering over other clustering strategies, e.g., k-means, is that a a generative model can generate random instances by estimation of data density.

The most representative generative model is the Gaussian mixture model (GMM) (Reynolds, 2009), which assumes that the observed data instances are generated from a mixture of Gaussian distributions. Though GMM has successful applications in many domains, its lack of deep structure could limit its capability to capture and represent the nonlinearity and complexity in some data types and thus adversely affect its clustering performance in many circumstances. To tackle this problem, deep generative models have been proposed to integrate the classical generative models with the neural networks, which have sufficient representation power to model the nonlinearity and complexity in the data. Deep generative models have many advantages, such as good explainability, high flexibility, and capability to generate new data instances (Zhou et al., 2022). This study attempts to apply both shallow generative clustering based on mixture models and deep generative clustering based on variational auto-encoder (VAE) (Kingma and Welling, 2013; Rezende, Mohamed, and Wierstra, 2014).

The objective of this study is fourfold. First, the study will identify how many robust economic preference types there are in the population (i.e., the optimal number of clusters) and the relative proportions of each preference type in the experimental sample. The empirical results could provide supporting evidence for the theory of preference heterogeneity. Second, the clustering results can describe what these preference types look like (i.e., where the cluster centroids are). These portraits might provide insights into studies about the long-standing controversial interplay and co-evolution of economic preferences. Theories suggest that preferences are supposed to be intertwined. For example, risk and time are intertwined because the present is deterministic while the future is inherently unknown and risky. Risk and ambiguity are difficult to distinguish as both are regarding uncertain outcomes in the future. Third, the clustering results will provide the probabilistic classification from the soft assignment of each individual to one of the preference types. This generative approach allows us to make out-of-sample predictions even if we cannot estimate the shape of one individual's utility function.

Finally, I would test, compare and evaluate the performance of different generative clustering algorithms without the knowledge of ground-truth class labels. Therefore, this study is also a case study to solve a real-world problem of low-dimensionality and moderate sample size with generative clustering algorithms. Identifying preference types with clustering algorithms is not a trivial task, and even state-of-the-art clustering methods would not guarantee robust clustering results. In studying the pattern of behavioral regularity correlations, Chapman et al. (2018) find that the k-means family fails to produce additional insight as the algorithms produce an optimal number of clusters of one, and they hypothesize that the "curse of dimensionality" problem prevents the identification of more subtle clusters. Gerlach et al. (2018) develop a data-driven approach to the characterizations of personality trait types, but they find that GMM will likely produce spurious clusters even if they carefully examine the results from model selection criteria, such as Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC).

2 Literature Review

2.1 Preference heterogeneity and clustering

Mixture modeling is one of the most popular methods to model preference heterogeneity and homogeneity in economic and decision science literature (Allenby and Rossi, 1998). Ferecatu and Önçüler (2016) use a mixture model to capture the heterogeneity in time discounting. The mixture model suggests that half of the subjects behave as standard exponential discounters, but the remaining half behave like hyperbolic discounters. Bruhin, Fehr, and Schunk (2019) use a mixture model to model the heterogeneity in social preference. They suggest that finite mixture models can endogenously identify the various preference types in the population without any presumptions regarding the existence or the properties of preference portraits. Bruhin, Fehr-Duda, and Epper (2010), Fehr-Duda, Epper et al. (2012), and Brown and Kim (2014) use mixture models to capture the heterogeneity in subjects' risk attitudes and group those individuals according to their degrees of risk aversion.

Other clustering algorithms have also been used in exploring similar problems. In particular, partition-based clustering (i.e., k-means family) is the most popular one due to its high explainability and tractability. Ferecatu and Önçüler (2016) examine the degree of heterogeneity in risk aversion and patience in the study sample, and they use k-means to classify the experimental subjects according to their individual-level estimates of risk aversion and discount rates. Chowdhury, Sutter, and Zimmermann (2022) use clustering analysis to group rural families into one of two clusters, with either relatively pro-social, risk-seeking, and patient members or relatively anti-social, risk averse, and impulsive family members. The clustering method they use is k-medoids. k-medoids is a variant of k-means and is more robust to outliers and noise than the latter. There are also challenges and caveats of blindly applying clustering algorithms to identify preference types. Chapman et al. (2018) apply k-medians to a large number of behavioral regularities to study the pattern of behavioral regularity correlations. However, they find that these methods fail to produce additional insight as the algorithms produce an optimal number of clusters of one, and they hypothesize that the "curse of dimensionality" problem prevents the identification of more subtle clusters.

2.2 Mixture models and clustering

A mixture model is a probabilistic method to represent and model the existence of subgroups in the entire population (McLachlan and Basford, 1988). Mixture models can be used for clustering tasks by outputting soft assignments of data instances. The most representative mixture model is GMM (Reynolds, 2009), which assumes that the data instances are generated from a mixture of Gaussian distributions. The expectation-maximization (EM) algorithm proposed by Dempster, Laird, and Rubin (1977) can be employed to learn the optimal parameters and soft assignments, whose convergence properties were later proved by Wu (1983). In the formulation of the EM algorithm, the updated set of parameters in M-step is chosen to maximize the data log-likelihood. While this ensures the greatest increase in likelihood, it is however possible to relax the requirement of maximization to simply increasing data log-likelihood. This approach is known as the generalized expectation maximization (GEM) algorithm and is often useful in cases where optimization of parameters during the M-step is difficult (Borman, 2004).

Although the most common choices of mixtures are Gaussian and categorical (for discrete features), other candidates for the distribution of the mixture components are available, for example, gamma distribution in univariate cases (Webb, 2000; Wiper, Insua, and Ruggeri, 2001; Liu et al., 2019) and Wishart distribution in multivariate cases (Carvalho, Sant'Anna Bins, and Siqueira Sant'Anna, 2019; Hidot and Saint-Jean, 2010; Liu et al., 2018; Yang et al., 2016). Clustering with mixture models has great potential for applications in many fields, such as computer vision (Zivkovic, 2004), speaker recognition (Reynolds, Quatieri, and Dunn, 2000), bioinformatics (Medvedovic and Sivaganesan, 2002; Lartillot and Philippe, 2004), financial analysis (Brown and Kim, 2014), and marketing analytics (Allenby and Rossi, 1998).

It is necessary a priori to limit the number of mixture components to be finite in a finite GMM. Thus Rasmussen (1999) proposes an infinite GMM that avoids the tricky procedure of deciding the appropriate number of mixture components. Bayesian methods of GMM clustering can decide the optimal number of mixture components automatically (Svensén and Bishop, 2005; Constantinopoulos, Titsias, and Likas, 2006); therefore, the difficult task of adjusting model complexity disappears, and the overfitting problem is circumvented. Roberts et al. (1998) show that the Bayesian GMM approach can select an optimal number of components in the mixture model, and its performance is comparable to other strategies of optimal model selection. For Bayesian GMM, variational Bayesian methods are used to derive an analytical approximation to the posterior of the latent features for statistical inference. Variational inference is also used for performing model specification selection by deriving a lower bound for the marginal likelihood (i.e., the evidence lower bound) of the observed data (Corduneanu and Bishop, 2001; Ghahramani and Beal, 1999). Variational inference for Bayesian GMM can be regarded as an extension of the classical EM algorithm used in GMM.

2.3 Mixture models and deep clustering

Classical clustering methods assume data are represented as features in a vectorized form through representation learning. As the data inputs become increasingly complex and high-dimensional, the traditional methods can no longer handle the challenging data types. To tackle the growing challenges in the clustering tasks, one strand of studies extend the usage of deep neural networks from classification focus to learning a clustering representation, which could result in a notable increase in clustering performance (see the comprehensive surveys of Zhou et al. (2022), Min et al. (2018), and Aljalbout et al. (2018)). This literature develops the concept of "deep clustering", which jointly optimizes the representation learning and clustering.

Zhou et al. (2022) classify those deep clustering methods developed in recent years into four streams based on the interplay between the representation learning module and the clustering module: (1) multi-stage (i.e., representation learning is sequentially linked with clustering), (2) iterative (i.e., representation learning and clustering are iteratively updated), (3) simultaneous (i.e., representation learning and clustering are simultaneously updated), and (4) generative (i.e., clustering is modeled as a prior representation).

Though GMM has successful applications in many domains in academia and industry, its lack of deep structure could limit its capability to capture and represent the nonlinearity and complexity in some data types and thus adversely affect its clustering performance in many circumstances. To tackle this problem, deep generative models have been proposed to integrate the classical generative models with the neural networks, which have sufficient representation power to model the nonlinearity and complexity in the data. There are primarily two types of generative model-based deep clustering methods: the methods based on variational auto-encoder (VAE) (Kingma and Welling, 2013; Rezende, Mohamed, and Wierstra, 2014) and the methods based on generative adversarial networks (GAN) (Goodfellow et al., 2014). Deep generative models have many advantages, such as good explainability, high flexibility, and capability to generate new data instances (Zhou et al., 2022). The generative clustering models could inherit the merits of neural network representation modules if they are systematically integrated.

VAE can be regarded as a generative variant of the well-known auto-encoder (AE), and it has a different mathematical formulation from AE as the former requires the latent space of AE to be a pre-specified mixture of distributions rather than a fixed vector. VAE is one of the variational Bayesian methods and inherits the merits of high scalability and flexibility of neural networks. It integrates the idea of neural networks to fit the conditional posterior and therefore, can optimize the objective via the variants of stochastic gradient descent (SGD) and backpropagation (BP) algorithm for neural networks (Min et al., 2018). For the clustering tasks of high-dimensional and complex data, one promising solution is to stack GMM with a deep neural network directly. GMM generates a latent vector, and the deep neural network further transforms the latent vector into complex data instances so that the integrated model can enjoy the merit of the latent cluster structure and has sufficient capacity to model the complex data. The two pioneering and representative algorithms in the deep generative clustering literature are VaDE (Jiang et al., 2017) and GMVAE (Dilokthanakul et al., 2016). More improved variants of VAE-based generative deep clustering methods have been proposed in recent years, such as (Prasad, Das, and Bhowmick, 2020; Li et al., 2020; Ji et al., 2021; Wang, Bao, and Guo, 2022; Yang et al., 2019).

Recent years have witnessed the notable success of GAN applications in the field of computer vision. GAN framework consists of a zero-sum adversarial game between two components of network: one network component is called “generator” and the other is called “discriminator”. The generator’s target is to synthesize fake and actual data instances to fool the discriminator by generating better instances, while the discriminator’s target is to discriminate the actual instances from the fake instances by estimating the probability that one new input is an actual instance taken from the data distribution. The generator network and the discriminator network can be optimized alternatively using variants of SGD. With the development of an adversary game between the two GAN components, the generator can finally output fake instances that follow a similar probability distribution of the observed data (Min et al., 2018). Considering its intriguing idea and outstanding performance, many studies are inspired to integrate GAN into generative clustering models. Specifically, Ben-Yosef and Weinshall (2018) proposed a simple framework to integrate a GMM with a GAN, in which GMM is a prior distribution for generating data instances. There are also some improved variants of GAN-based deep generative clustering algorithms, such as (Yu and Zhou, 2018; Ntelemis, Jin, and Thomas, 2021; Mukherjee et al., 2019; Gan et al., 2021; Ghasedi et al., 2019).

Although deep generative clustering models can generate new instances while completing clustering, they also have some weaknesses. Training a generative model usually involves Monte Carlo sampling, which may incur training unstable and high computational complexity. The mainstream generative models are based on VAE and GAN, and inevitably inherit the same disadvantages of them. VAE-based models usually require prior assumptions on the data distributions, which may not be held in real cases. Although GAN-based algorithms are more flexible and diverse, they usually suffer from mode collapse and slow convergence, especially for the data with multiply clusters (Zhou et al., 2022).

3 Data

The Global Preferences Survey is a globally representative dataset on risk and time preferences, positive and negative reciprocity, altruism, and trust. We collected these preference data as well as a rich set of covariates for 80,000 individuals, drawn as representative samples from 76 countries around the world, representing 90 percent of both the world’s population and global income. The dataset is owned by Armin Falk (briq). The preference module is a concise, experimentally-validated survey module for measuring risk aversion, time discounting, trust, altruism, positive and negative reciprocity. These preferences affect individuals’ choices in myriad situations. The module is a convenient tool for obtaining standardized measures in all popular methods of data collection. It is therefore a useful tool in a wide range of applications, not least because preference measures can allow for improved prediction of many important economic behaviors, or can provide control variables if researchers want to identify causal effects of other factors that are correlated with preferences.

Time Preference/Patience. The measure of time preference is derived from the combination of responses to two survey measures, one with a quantitative and one with a qualitative format. The quantitative survey measure consists of a series of five interdependent hypothetical binary choices between immediate and delayed financial rewards. In each of the five questions, participants had to decide between receiving a payment today or larger payments in 12 months. The qualitative measure of patience is given by the respondents' self-assessment regarding their willingness to wait on an 11-point Likert scale, asking "how willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?"

Risk Preference. Risk preferences were elicited through a series of related quantitative questions as well as one qualitative question. Just as with patience, the quantitative measure consists of a series of five binary choices. Choices were between a fixed lottery, in which the individual could win x or zero, and varying sure payments, y. Choice of the lottery resulted in an increase of the sure amount being offered in the next question, and vice versa, thereby zooming in around the individual's certainty equivalent. The qualitative item and the outcome of the quantitative staircase measure were combined through roughly equal weights.

Positive Reciprocity. Respondents' propensities to act in a positively reciprocal way were measured using one quantitative item and one qualitative question. First, respondents were presented a choice scenario in which they were asked to imagine that they got lost in an unfamiliar area and that a stranger – when asked for directions – offered to take them to their destination. Respondents were then asked which out of six presents (worth between 5 and 30 euros, or the respective country-specific equivalents) they would give to the stranger as a "thank you". Second, respondents were asked to provide a self-assessment about how willing they are to return a favor on an 11-point Likert scale. These two items receive roughly equal weights.

Negative Reciprocity. Negative reciprocity was elicited through three self-assessments. First, respondents were asked how willing they are to take revenge if they are treated very unjustly, even if doing so comes at a cost (Likert scale, 0-10). The second and third item probed respondents about their willingness to punish someone for unfair behavior, either towards themselves or towards a third person. This last item captures prosocial punishment and hence a concept akin to norm enforcement. These three items receive weights of about one third each.

Altruism. Altruism was measured through a combination of one qualitative and one quantitative item, both of which are related to donations. The qualitative question asked respondents how willing they would be to give to good causes without expecting anything in return on an

11-point scale. The quantitative scenario depicted a situation in which the respondent unexpectedly received 1,000 euros and asked them to state how much of this amount they would donate. These two items were weighted about equally.

Trust. The trust measure is based on one item, which asked respondents whether they assume that other people only have the best intentions (Likert scale, 0-10). The item was a strong predictor of trusting behavior in incentivized trust games, in the survey design stage. Time constraints and the fact that there already exists a global measure of trust in the World Values Survey (WVS) data set determined the choice to have only one item measuring trust.

4 Methodology and Algorithms

4.1 Preprocessing and pretest

Preference features have different scales as decision task responses have different score ranges. Before clustering, the data are preprocessed with standardization by removing the mean and scaling to unit variance. Kernel principal component analysis (PCA) is used to extract the preference features used in clustering analysis. Kernel PCA is an extension of classical linear PCA using kernel methods. With a pre-specified kernel function, the operations of linear PCA are performed in a reproducing kernel Hilbert space (RKHS) so that kernel PCA can capture the nonlinear properties in the original features. After reducing dimensionality with kernel PCA (with a polynomial kernel of degree 2), four dimensions are left, each corresponding to one dimension of the preferences. Figure 1 plots the individual and cumulative explained variance ratio as a function of the number of principal components for each preference. Except for the risk preference, the first principal component can explain about half of the variance of responses to one specific group of decision tasks.

To examine to what degree separable clusters exist in the preprocessed data, a cluster tendency test is performed before clustering (Adolfsson, Ackerman, and Brownstein, 2019). I use Hopkins statistic (Hopkins and Skellam, 1954), a hypothesis test statistic with the null hypothesis that the data are uniformly randomly distributed. If the test is positive (Hopkins score close to 0, for the implementation of *pyclustertend*), it suggests that the data is not uniformly distributed, and clustering can be helpful to classify the data instances; if the statistic is too high (Hopkins score close to 0.5), then clustering is not the suitable technique for studying the problem. The preliminary result of the Hopkins score is around 0.05 with a sampling rate of 10 percent, suggesting strong clusterability in the survey data.

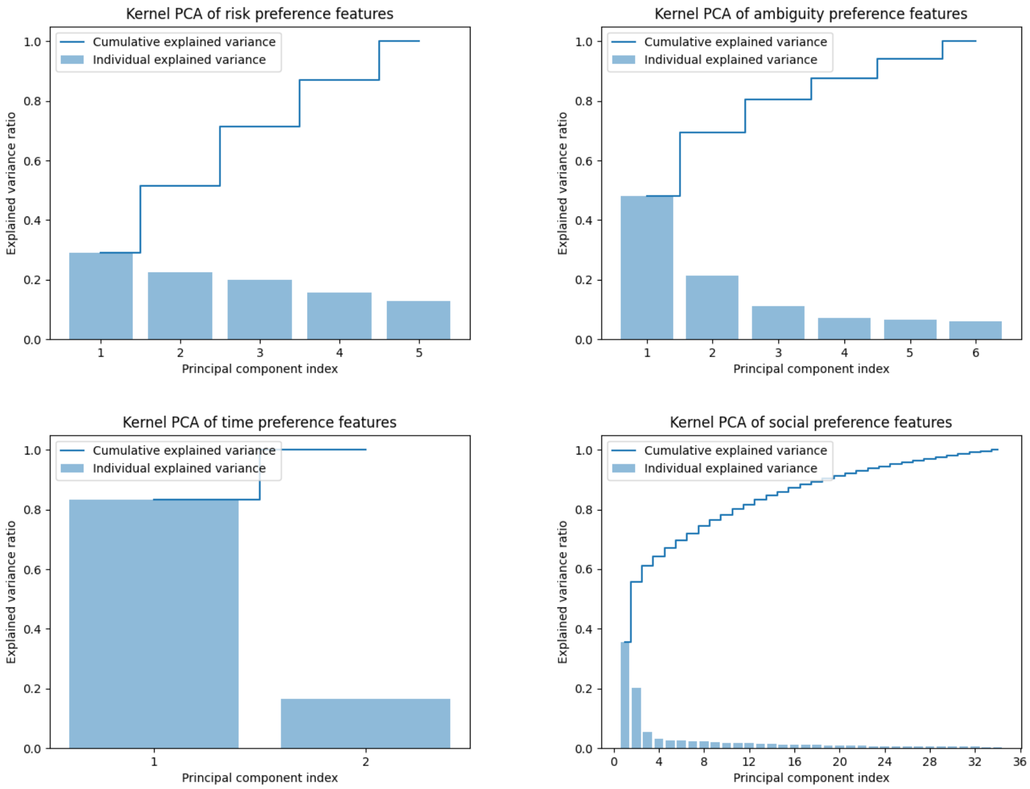

Figure 1: No. of principal components and explained variance ratio

4.2 Generative model-based shallow clustering

This study will first use a mixture model-based generative method clustering to identify robust economic preference types with the survey data. Specifically, a GMM will be used as the baseline algorithm for comparison with other candidate clustering methods. To use GMM, it is essential to determine the optimal number of clusters before clustering. The choice is not guided by any prior knowledge of this problem, and it will also affect the centroids of the clusters. The first method to determine the optimal number of clusters is the information criterion approach for finite mixture models, i.e., Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC). An exhaustive search of the number of clusters and GMM covariance matrice types based on BIC values suggests that a GMM with diagonal covariance matrices of 12 components has the lowest BIC value.

The second method to determine the optimal number of clusters is to use the Bayesian GMM (BGMM) with the Dirichlet Process (DP) as a prior distribution on the number of clusters. Bayesian methods of clustering can automatically decide the optimal number of mixture components (Svensén and Bishop, 2005; Constantinopoulos, Titsias, and Likas, 2006); therefore, the difficult task of adjusting model complexity disappears, and the overfitting problem is avoided. Roberts et al. (1998) show that the Bayesian GMM approach can select an optimal number of components in the mixture model, and its performance is nearly comparable to other strategies of optimal model selection. The effective number of components in BGMM can be computed from the data themselves. With a fixed total number of 20 of the Gaussian components and a component weight threshold of 0.01, the analysis suggests that the number of effective Gaussian components is also 12, which is consistent with the results of GMM.[1]

Grandmont (1987) theoretically proves that, under suitable regularity assumptions, the class of distributions of agents' preference is the gamma distribution. Weitzman (2001) and Silver, Slud, and Takamoto (2002) also suggest that the stochastic preferences of agents under certain conditions follow gamma distributions by providing empirical and simulation evidence. Therefore, for a univariate case, a gamma mixture model might be preferred over GMM and used as the second model. For the multivariate case, the multivariate counterpart of gamma distribution should be used as the underlying distribution of the mixture model. Wishart distribution is a generalization to multiple dimensions of the gamma distribution. For a p-dimensional random variable X following a Wishart distribution, i.e., $X \sim W_p(V, n)$, the parameter $n > p - 1$ is the degrees of freedom, and the $p \times p$ positive definite matrix $V > 0$ is the scale matrix of the distribution. The Wishart distribution could be used for visualizing distributions of the covariance matrices (Tokuda et al., 2011). The Wishart mixture models (WMM) have been proposed for modeling and clustering purposes in fields of movement clustering (Hidot and Saint-Jean, 2010) and radar image analysis (Yang et al., 2016; Liu et al., 2018; Carvalho, Sant'Anna Bins, and Siqueira Sant'Anna, 2019). In this project, I implement a WMM as the third clustering algorithm for evaluation and comparison. In the standard formulation of the EM-GMM algorithm, the updated set of parameters θ_{n+1} in step $(n+1)$ in M-step should be chosen as the value of θ such that $L(\theta) - L(\theta_n) \triangleq \Delta(\theta|\theta_n)$ is maximized. This optimization can ensure the greatest increase in likelihood $L(\theta)$ and high convergence speed in EM-GMM. For WMM, it is difficult to maximize during the M-step due to the complex form of probability distribution function (PDF) of Wishart distribution, so I have to relax the requirement of maximization to simply increasing $\Delta(\theta|\theta_n)$ so that $\Delta(\theta_{n+1}|\theta_n) \geq \Delta(\theta_n|\theta_n)$. This method of just increasing but not necessarily maximizing data log-likelihood is called the generalized expectation maximization (GEM) algorithm and is applicable in many situations where the optimization during M-step is difficult (Borman, 2004).

[1] Some alternative methods to determine the optimal number of clusters include the visualization approach (e.g., the elbow method) and the clustering stability method. For clustering stability, one should choose the number of clusters so that the clustering algorithm can produce stable partitions when the data are repeatedly perturbed (Von Luxburg et al., 2010).

4.3 Generative model-based deep clustering

Though GMM has successful applications in many domains, both in academia and industry, its lack of deep structure could limit its capability to capture and represent the nonlinearity and complexity in some data types and thus adversely affect its clustering performance in many circumstances. To tackle this problem, deep generative models have been proposed to integrate the classical generative models with the neural networks developed in deep learning studies, which have sufficient representation power to model the nonlinearity and complexity in the data. As mentioned in the review of related works, there are primarily two types of generative model-based deep clustering methods: the methods based on variational auto-encoder (VAE) (Kingma and Welling, 2013; Rezende, Mohamed, and Wierstra, 2014) and the methods based on generative adversarial networks (GAN) (Goodfellow et al., 2014). Deep generative models have many advantages, such as good explainability, high flexibility, and capability to generate new data instances. Integrating the classical generative clustering models with deep learning representation could be promising and insightful (Zhou et al., 2022). This study will also attempt to apply VAE-based deep clustering to this problem.

In particular, this project attempts to adopt the architecture of Jiang et al. (2017)'s work. They propose a novel unsupervised generative clustering algorithm named Variational Deep Embedding (VaDE) based on the classical neural network framework of VAE. Specifically, VaDE models the data generative procedure with a GMM and a deep neural network (DNN). First, the GMM picks a cluster from which a latent embedding is generated. Then the DNN decodes the latent embedding into an observable. They compare their method with many baselines methods, such as GMM, GMM+VAE, and GMM+AE. Their experimental results prove that VaDE outperforms the state-of-the-art clustering techniques on five benchmark tasks, including both image processing and natural language processing. The most relevant work of Jiang et al. (2017)'s VaDE is the stacked generative semi-supervised model developed by (Kingma et al., 2014). The major difference between their approaches is that the latter applies their model to semi-supervised classification tasks, whereas Jiang et al. (2017) focus on unsupervised clustering. Variational Deep Embedding (VaDE) is a model for probabilistic clustering problem within the framework of VAE as described by Jiang et al. (2017).

Jiang et al. (2017)'s experimental results show that VaDE suffers from the same problem as described in Kingma et al. (2016). The reconstruction term is likely to be not strong enough at the beginning of training, and thus, the model is likely to be stuck in some undesirable local optimal or saddle point. Their solution is to use a pretraining strategy to tackle this problem. They propose a stacked autoencoder to pretrain the neural networks. Then all data instances

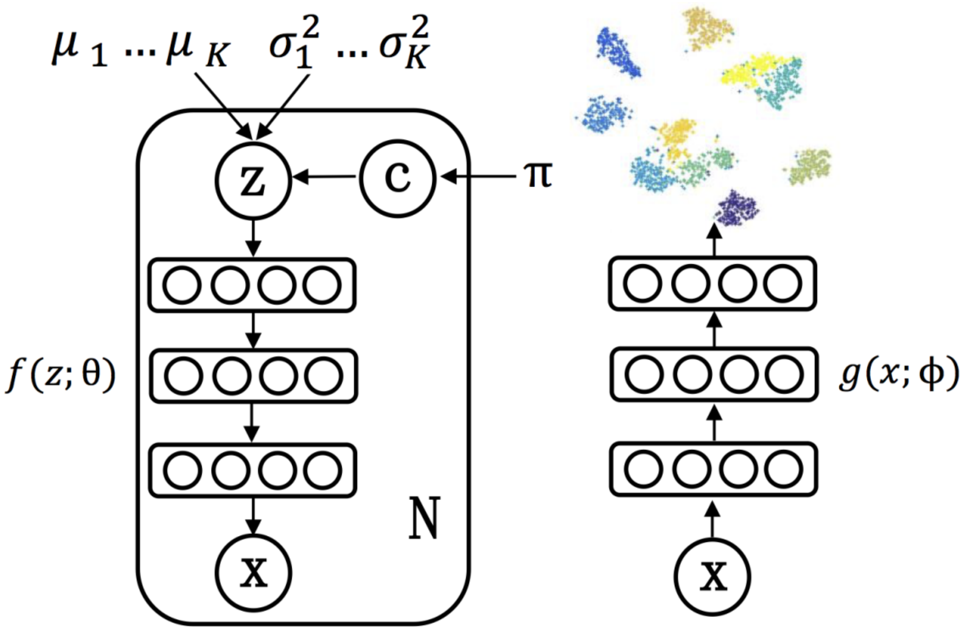

Figure 2: The diagram of VaDE, adapted from Jiang et al. (2017). The data generative process of VaDE is done as follows: (1) a cluster is picked from a GMM model; (2) a latent embedding is generated based on the picked cluster; (3) DNN $f(\mathbf{z};\theta)$ decodes the latent embedding into an observable $\mathbf{x}$. An encoder network $g(\mathbf{x};\phi)$ is used to maximize the ELBO of VaDE.

are projected into the latent space by the pretrained decoder, where a standard GMM is used for parameter initialization for the encoder and decoder. They show that the pretraining strategy works well, and a good initialization is achieved with a small number of epochs in VaDE. They also report that VaDE is not sensitive to hyperparameters after pretraining.

In this project, the VaDE network architecture and experimental setup are revised for this specific clustering task: (1) Number of epochs is how many times the model will go through the dataset to update the neural network model parameters during training. In this project, the number of epochs is set to 50 (and 20 in pretraining). (2) Batch size is the number of instances involved in updating the neural network model in a training loop. In this project, considering the sample size is small, the batch size is set to 4. (3) Learning rate controls the convergence speed of the neural network model to reach the final solution. If the learning rate is too low, it requires more epochs to train the neural network model. While if it is too high, the learning process may be stuck in a local optimum rather than reaching the global optimal solution. In this project, the learning rate is set to 0.001. The learning rate also reduces after every ten epochs with a decay rate of 0.9. (4) Number of hidden layer proxies how "deep" one neural network model is. Unlike the tasks in natural language processing or computer vision, the task in this project is not complex; therefore, three hidden layers for both the encoder and decoder are "deep" enough to capture the nonlinearity in the features. (5) Number of neurons in layers controls the size of the neural network. In this project, considering the task has a dimension of four, the architectures of the encoder f and the decoder g are designed as $k = 12 - 16 - 4 - 4 - d = 4$ and $d = 4 - 4 - 4 - 16 - k = 12$, respectively, where $k = 12$ is the number of clusters and $d = 4$ is the input dimensionality. All layers in the network are densely connected layers. (6) Activation function controls how the input is transformed into a nonlinear output in the network. In this project, the popular function ReLU is used for each node. (7) Loss function is used to supervise the model training by minimizing the "distance" between the model predictions and the desired outputs (e.g., ground-truth labels). Usually, the binary cross-entropy (BCE) is used for classification tasks; however, using BCE leads to an error reported by *pytorch*; therefore, this project uses mean squared error (MSE) instead as the loss function. (8) Optimizer is the algorithm used to update the weights of neural networks in training with the objective of minimizing the loss function. In this project, Adaptive Moment Estimation (Adam) optimizer (Kingma and Ba, 2015) is used to maximize the ELBO.

5 Performance Analysis and Evaluation

5.1 Performance evaluation and comparison

Without the knowledge of the ground-truth class labels, internal criteria should be used to evaluate clustering quality (Arbelaitz et al., 2013). One popular metric to evaluate the clustering performance is Calinski-Harabasz index (C-H index). C-H index calculates the ratio of two sums for all clusters. The first is the sum of between-clusters dispersion, and the second is the sum of within-cluster dispersion. The dispersion is calculated as the sum of squared distances. The index value is higher when clusters are dense and properly partitioned, consistent with the standard concept of a cluster. Table 1 reports the (1) data log-likelihood per sample, (2) BIC, and (3) Calinski-Harabasz index for each of the four candidate models aforementioned. First, among these four clustering models, GMM has the highest data log-likelihood per sample, followed by BGMM. While WMM has relatively low data log-likelihood. This result suggests that GMM fits the data better, and its parameters estimated by the EM algorithm explain the observed data better than other models. Second, BIC prefers GMM over WMM as the GMM has a lower BIC value. BIC does not apply to BGMM and VaDE due to the variational inference. Third, GMM also has the highest Calinski-Harabasz index value, followed by BGMM. This suggests that clustering results are dense and well partitioned by GMM and BGMM in terms of between-clusters dispersion and within-cluster dispersion. These results prove that GMM, as one of the most widely used generative clustering algorithms, has outstanding performance over other models. In subsequent analysis, I use the results of GMM as the baseline for preliminary performance analysis.

Table 1: Clustering model selection and performance evaluation metrics

Metrics	GMM	BGMM	WMM	VaDE
Data Log-likelihood Per Sample	-3.14	-3.48	-13.81	N.A.
Bayesian Information Criterion (BIC)	20522.3	N.A.	87898.14	N.A.
Calinski Harabasz Index	176.61	164.59	43.72	93.23

T-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten and Hinton, 2008) is a machine learning technique used for visualizing high-dimensional data through projecting each data instance into 2D or 3D. Unlike linear PCA, t-SNE is a nonlinear dimensionality reduction technique designed for embedding high-dimensional data for visualization in a low-dimensional space of 2D or 3D. In particular, t-SNE projects each high-dimensional items into a 2D or 3D points in the lower-dimensional space. The method models similar items by their nearest neighbors and models dissimilar items by their distant neighbors with high probability. The parameter of perplexity in t-SNE controls the number of nearest neighbors that are used, and different values of perplexity can lead to significantly different results. I test different values of perplexity for the clustering results. Figure 3 reports t-SNE 2D visualization of the baseline GMM clustering results (with 12 Gaussian components and diagonal covariance matrices). The t-SNE visualization results suggest the task of identifying economic preference types is not a trivial task as a large proportion of modeled points are overlapped even for GMM clustering. The relatively stable and unoverlapped clusters in the four subfigures are the clusters of label 2 (i.e., yellow points) and label 10 (i.e., purple points).

5.2 Cluster weights and centroids

Table 2 report the cluster weights and centroids of the 12 robust clusters based on the baseline GMM clustering results (with 12 Gaussian components and diagonal covariance matrices). As the features (i.e., the first principal components for each preference in PCA) are standardized to zero mean and unit variance, the relative values of cluster centroids can be used to describe the portraits and characteristics of the clusters. The largest cluster of economic preferences accounts for about 23 percent of the population. The subjects in the largest cluster, as portrayed by their cluster centroid, are relatively risk-seeking, relatively ambiguity-seeking, very patient and relatively anti-social. The second largest cluster accounts for about 18 percent of the population.

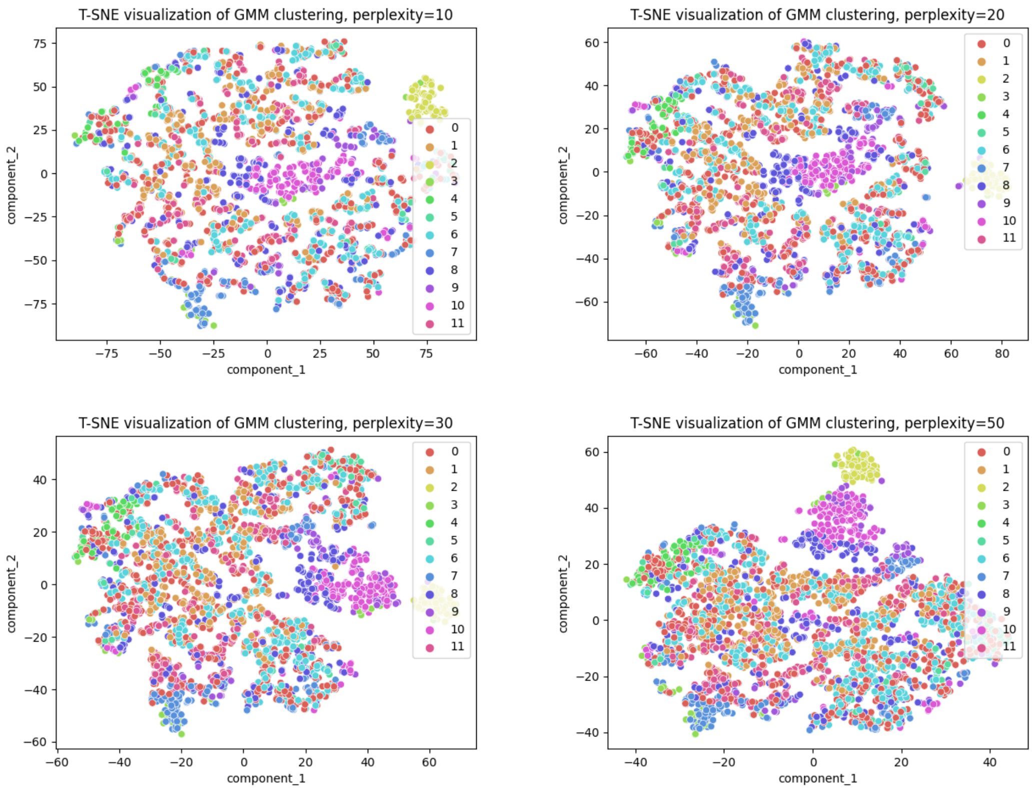

Figure 3: t-SNE visualization of GMM clustering results (12 Gaussian components and diagonal covariance matrices)

The subjects in the second largest cluster are relatively risk averse, relatively ambiguity averse, very patient, and relatively pro-social. The third largest cluster accounts for about 14 percent of the population. The subjects in the third largest cluster are very risk-seeking, relatively ambiguity-seeking, very patient, and very pro-social. The last cluster accounts for about 1 percent of the population. They are described as very risk averse, very ambiguity averse, very impatient, and relatively pro-social.

On the one hand, there are 12 robust clusters in the population each accounting for more than 1 percent of the population. This result is consistent with the theory of preference heterogeneity in the population. On the other hand, the first three largest clusters account for about 55 percent of the population. This result suggests the universal phenomenon of preference homogeneity in the population. One economic theory suggests that risk and time are intertwined because the present is deterministic while the future is inherently unknown and risky. While the clustering results do not support this theory as the values of “risk-seeking” and “impatience” of the cluster centroids are not always of the same sign. Another economic theory suggests that risk and ambiguity are difficult to distinguish as both are regarding uncertain outcomes in the future. There is some supportive evidence for this theory as the values of “risk-seeking” and “ambiguity-seeking” of the cluster centroids are usually of the same sign.

Table 2: Cluster weights and centroids of GMM clustering results (12 Gaussian components and diagonal covariance matrices)

Type	Cluster Weight	Cluster Centroid		
		Risk Seeking	Impatience	Pro-sociality
1	0.23	0.05	-0.39	-0.13
2	0.18	-0.03	-0.42	0.15
3	0.14	0.44	-0.43	0.46
4	0.08	0.02	0.2	-0.13
5	0.08	-0.07	1.21	0.07
6	0.08	-0.11	-0.25	0.03
7	0.07	0.18	-0.24	0.04
8	0.04	-0.39	-0.41	-1.71
9	0.03	-1.32	-0.39	-0.2
10	0.03	0.03	0.59	0.03
11	0.03	-0.32	4.18	0.13
12	0.01	-0.84	4.18	0.03

6 Conclusions and Future Work

This study attempts to use clustering analysis to identify robust economic preference types of people with a comprehensive task-level dataset from Global Preference Survey. Specifically, this study uses generative model shallow clustering methods (i.e., GMM, BGMM, WMM) and deep clustering method (i.e., VaDE) to identify robust economic preference types. The results suggest that there are about 12 diverse and robust preference types in the dataset, with the top three types accounting for about 55 percent of the subjects. The classical GMM is still the most robust the effective algorithm for this specific clustering task, while other clustering techniques also show great potential in clustering capability in other tasks.

The primary direction of future work is to design new models or architectures for this specific task using the domain knowledge of economic preference theory. Another possible extension of the current study is to use clustering ensembles techniques (Ghaemi et al., 2009). The clustering ensembles synthesize different clustering outputs generated by different clustering models into a final clustering output (Ghaemi et al., 2009). Clustering ensemble is a promising technique for increasing accuracy, robustness, and stability of the final clustering partitions. Finally, one limitation that should be addressed in future work is subsample clustering analysis by partitioning the entire survey dataset into subsets according to population sources (Beijing subjects vs. Singapore subjects), gender (male subjects vs. female subjects), and sessions (one primary and one replication session in each city).

7 Bibliography

Adolfsson, Andreas, Margareta Ackerman, and Naomi C Brownstein. 2019. "To cluster, or not to cluster: An analysis of clusterability methods." *Pattern Recognition* 88:13–26.

Aljalbout, Elie, Vladimir Golkov, Yawar Siddiqui, Maximilian Strobel, and Daniel Cremers. 2018. "Clustering with deep learning: Taxonomy and new methods." *arXiv preprint arXiv:1801.07648* .

Allenby, Greg M and Peter E Rossi. 1998. "Marketing models of consumer heterogeneity." *Journal of Econometrics* 89 (1-2):57–78.

Arbelaitz, Olatz, Ibai Gurrutxaga, Javier Muguerza, Jesús M Pérez, and Iñigo Perona. 2013. "An extensive comparative study of cluster validity indices." *Pattern Recognition* 46 (1):243–256.

Ben-Yosef, Matan and Daphna Weinshall. 2018. "Gaussian mixture generative adversarial networks for diverse datasets, and the unsupervised clustering of images." *arXiv preprint arXiv:1808.10356* .

Borman, Sean. 2004. "The expectation maximization algorithm - a short tutorial."

Brown, Alexander L and Hwagyun Kim. 2014. "Do individuals have preferences used in macro-finance models? An experimental investigation." *Management Science* 60 (4):939–958.

Bruhin, Adrian, Ernst Fehr, and Daniel Schunk. 2019. "The many faces of human sociality: Uncovering the distribution and stability of social preferences." *Journal of the European Economic Association* 17 (4):1025–1069.

Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper. 2010. "Risk and rationality: Uncovering heterogeneity in probability distortion." *Econometrica* 78 (4):1375–1412.

Carvalho, Naiallen Carolyne Rodrigues Lima, Leonardo Sant'Anna Bins, and Sidnei João Siqueira Sant'Anna. 2019. "Analysis of stochastic distances and Wishart mixture models applied on PolSAR images." *Remote Sensing* 11 (24):2994.

Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer. 2018. "Econographics." Tech. rep., National Bureau of Economic Research.

Chowdhury, Shyamal, Matthias Sutter, and Klaus F Zimmermann. 2022. "Economic preferences across generations and family clusters: A large-scale experiment in a developing country." *Journal of Political Economy* .

Constantinopoulos, Constantinos, Michalis K Titsias, and Aristidis Likas. 2006. "Bayesian feature and model selection for Gaussian mixture models." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (6):1013–1018.

Corduneanu, Adrian and Christopher M Bishop. 2001. "Variational Bayesian model selection for mixture distributions." In *Artificial Intelligence and Statistics*, vol. 2001. Morgan Kaufmann Waltham, MA, 27–34.

Dempster, Arthur P, Nan M Laird, and Donald B Rubin. 1977. "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society: Series B (Methodological)* 39 (1):1–22.

Dilokthanakul, Nat, Pedro AM Mediano, Marta Garnelo, Matthew CH Lee, Hugh Salimbeni, Kai Arulkumaran, and Murray Shanahan. 2016. "Deep unsupervised clustering with Gaussian mixture variational autoencoders." *arXiv preprint arXiv:1611.02648* .

Fehr-Duda, Helga, Thomas Epper et al. 2012. "Probability and risk: Foundations and economic implications of probability-dependent risk preferences." *Annual Review of Economics* 4 (1):567–593.

Ferecatu, Alina and Ayse Önçüler. 2016. "Heterogeneous risk and time preferences." *Journal of Risk and Uncertainty* 53 (1):1–28.

Gan, Yanhai, Xinghui Dong, Huiyu Zhou, Feng Gao, and Junyu Dong. 2021. "Learning the precise feature for cluster assignment." *IEEE Transactions on Cybernetics* .

Gerlach, Martin, Beatrice Farb, William Revelle, and Luís A Nunes Amaral. 2018. "A robust data-driven approach identifies four personality types across four large data sets." *Nature Human Behaviour* 2 (10):735–742.

Ghaemi, Reza, Md Nasir Sulaiman, Hamidah Ibrahim, and Norwati Mustapha. 2009. "A survey: Clustering ensembles techniques." *International Journal of Computer and Information Engineering* 3 (2):365–374.

Ghahramani, Zoubin and Matthew Beal. 1999. "Variational inference for Bayesian mixtures of factor analysers." *Advances in Neural Information Processing Systems* 12.

Ghasedi, Kamran, Xiaoqian Wang, Cheng Deng, and Heng Huang. 2019. "Balanced self-paced learning for generative adversarial clustering network." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4391–4400.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. "Generative adversarial nets." *Advances in Neural Information Processing Systems* 27.

Grandmont, Jean-Michel. 1987. "Distributions of preferences and the "Law of Demand"." *Econometrica* :155–161.

Hidot, Sullivan and Christophe Saint-Jean. 2010. "An Expectation-Maximization algorithm for the Wishart mixture model: Application to movement clustering." *Pattern Recognition Letters* 31 (14):2318–2324.

Hopkins, Brian and John Gordon Skellam. 1954. "A new method for determining the type of distribution of plant individuals." *Annals of Botany* 18 (2):213–227.

Ji, Qiang, Yanfeng Sun, Junbin Gao, Yongli Hu, and Baocai Yin. 2021. "A decoder-free variational deep embedding for unsupervised clustering." *IEEE Transactions on Neural Networks and Learning Systems* .

Jiang, Zhuxi, Yin Zheng, Huachun Tan, Bangsheng Tang, and Hanning Zhou. 2017. "Variational deep embedding: An unsupervised and generative approach to clustering." *International Joint Conference on Artificial Intelligence (IJCAI-2017)* .

Kingma, Diederik P and Jimmy Ba. 2015. "Adam: A method for stochastic optimization." *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)* .

Kingma, Diederik P and Max Welling. 2013. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* .

Kingma, Durk P, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. 2014. "Semi-supervised learning with deep generative models." *Advances in Neural Information Processing Systems* 27.

Kingma, Durk P, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. 2016. "Improved variational inference with inverse autoregressive flow." *Advances in Neural Information Processing Systems* 29.

Lartillot, Nicolas and Hervé Philippe. 2004. "A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process." *Molecular Biology and Evolution* 21 (6):1095–1109.

Li, Zhihan, Youjian Zhao, Haowen Xu, Wenxiao Chen, Shangqing Xu, Yilin Li, and Dan Pei. 2020. "Unsupervised clustering through Gaussian mixture variational autoencoder with non-reparameterized variational inference and std annealing." In *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

Liu, Chi, Heng-Chao Li, Kun Fu, Fan Zhang, Mihai Datcu, and William J Emery. 2019. "Bayesian estimation of generalized Gamma mixture model based on variational EM algorithm." *Pattern Recognition* 87:269–284.

Liu, Chi, Wenzhi Liao, Heng-Chao Li, Kun Fu, and Wilfried Philips. 2018. "Unsupervised classification of multilook polarimetric SAR data using spatially variant Wishart mixture model with double constraints." *IEEE Transactions on Geoscience and Remote Sensing* 56 (10):5600–5613.

McLachlan, Geoffrey J and Kaye E Basford. 1988. *Mixture models: Inference and applications to clustering*, vol. 38. M. Dekker New York.

Medvedovic, Mario and Siva Sivaganesan. 2002. "Bayesian infinite mixture model based clustering of gene expression profiles." *Bioinformatics* 18 (9):1194–1206.

Min, Erxue, Xifeng Guo, Qiang Liu, Gen Zhang, Jianjing Cui, and Jun Long. 2018. "A survey of clustering with deep learning: From the perspective of network architecture." *IEEE Access* 6:39501–39514.

Mukherjee, Sudipto, Himanshu Asnani, Eugene Lin, and Sreeram Kannan. 2019. "Clustergan: Latent space clustering in generative adversarial networks." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33. 4610–4617.

Ntelemis, Foivos, Yaochu Jin, and Spencer A Thomas. 2021. "Image clustering using an augmented generative adversarial network and information maximization." *IEEE Transactions on Neural Networks and Learning Systems* .

Prasad, Vignesh, Dipanjan Das, and Brojeshwar Bhowmick. 2020. "Variational clustering: Leveraging variational autoencoders for image clustering." In *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–10.

Rasmussen, Carl. 1999. "The infinite Gaussian mixture model." *Advances in Neural Information Processing Systems* 12.

Reynolds, Douglas A. 2009. "Gaussian mixture models." *Encyclopedia of Biometrics* 741 (659-663).

Reynolds, Douglas A, Thomas F Quatieri, and Robert B Dunn. 2000. "Speaker verification using adapted Gaussian mixture models." *Digital Signal Processing* 10 (1-3):19–41.

Rezende, Danilo Jimenez, Shakir Mohamed, and Daan Wierstra. 2014. "Stochastic backpropagation and approximate inference in deep generative models." In *International Conference on Machine Learning*. PMLR, 1278–1286.

Roberts, Stephen J, Dirk Husmeier, Iead Rezek, and William Penny. 1998. "Bayesian approaches to Gaussian mixture modeling." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (11):1133–1142.

Silver, Jonathan, Eric Slud, and Keiji Takamoto. 2002. "Statistical equilibrium wealth distributions in an exchange economy with stochastic preferences." *Journal of Economic Theory* 106 (2):417–435.

Svensén, Markus and Christopher M Bishop. 2005. "Robust Bayesian mixture modelling." *Neurocomputing* 64:235–252.

Tokuda, Tomoki, Ben Goodrich, I Van Mechelen, Andrew Gelman, and F Tuerlinckx. 2011. "Visualizing distributions of covariance matrices." *Columbia Univ., New York, USA, Tech. Rep* :18–18.

Van der Maaten, Laurens and Geoffrey Hinton. 2008. "Visualizing data using t-SNE." *Journal of Machine Learning Research* 9 (11).

Von Luxburg, Ulrike et al. 2010. "Clustering stability: An overview." *Foundations and Trends in Machine Learning* 2 (3):235–274.

Wang, Wenqing, Junpeng Bao, and Siyao Guo. 2022. "Neural generative model for clustering by separating particularity and commonality." *Information Sciences* 589:813–826.

Webb, Andrew R. 2000. "Gamma mixture models for target recognition." *Pattern Recognition* 33 (12):2045–2054.

Weitzman, Martin L. 2001. "Gamma discounting." *American Economic Review* 91 (1):260–271.

Wiper, Michael, David Rios Insua, and Fabrizio Ruggeri. 2001. "Mixtures of gamma distributions with applications." *Journal of Computational and Graphical Statistics* 10 (3):440–454.

Wu, CF Jeff. 1983. "On the convergence properties of the EM algorithm." *The Annals of Statistics* :95–103.

Yang, Linxiao, Ngai-Man Cheung, Jiaying Li, and Jun Fang. 2019. "Deep clustering by Gaussian mixture variational autoencoders with graph embedding." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 6440–6449.

Yang, Wen, Xiangli Yang, Tianheng Yan, Hui Song, and Gui-Song Xia. 2016. "Region-based change detection for polarimetric SAR images using Wishart mixture models." *IEEE Transactions on Geoscience and Remote Sensing* 54 (11):6746–6756.

Yu, Yang and Wen-Ji Zhou. 2018. "Mixture of GANs for clustering." In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18)*. 3047–3053.

Zhou, Sheng, Hongjia Xu, Zhuonan Zheng, Jiawei Chen, Jiajun Bu, Jia Wu, Xin Wang, Wenwu Zhu, Martin Ester et al. 2022. "A comprehensive survey on deep clustering: Taxonomy, challenges, and future directions." *arXiv preprint arXiv:2206.07579* .

Zivkovic, Zoran. 2004. "Improved adaptive Gaussian mixture model for background subtraction." In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR-2004)*, vol. 2. IEEE, 28–31.