POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Executive Summary of the Thesis

Robust model-based clustering for high-dimensional data via covariance matrices regularization

Laurea Magistrale in Mathematical Engineering - Ingegneria Matematica

Authors: Luca Panzeri
Davide Zaltieri

Advisor: Dr. Andrea Cappozzo

Academic year: 2022-2023

The proliferation of contaminated high-dimensional datasets has been a defining trend in recent years. This surge in dimensionality and complexity poses a significant challenge for traditional clustering methods, as existing robust clustering methods suffer from the curse of dimensionality when p is large, while existing approaches for high-dimensional data are, in general, not robust. This thesis aims to address that challenge, by integrating high-dimensional covariance matrix estimators into the efficient TCLUST methodology for robust constrained clustering. While TCLUST has demonstrated strong performance in handling contaminated low-dimensional data, it faces two significant limitations when dealing with a substantial increase in the number of variables. The first limitation is related to inizialization. TCLUST relies on the initial random selection of only $k \times (p+1)$ observations to establish an outlier-free starting point. However, as the dimensionality p increases, this initial subset may exceed the total number of observations n, thereby increasing the risk of including outliers during the initialization phase. The second limitation arises from the parameters growth in the covariance matrices, causing them to become singular or ill-conditioned, resulting in a determinant equal to zero. Consequently, they become non-invertible and unmanageable using traditional methods reliant on matrix inversion. To address this challenge, regularization techniques are indispensable, as they introduce penalty terms within the covariance matrix estimation, allowing for reliable inference even when $p > n$.

1. TCLUST: a robust constrained clustering algorithm

Let $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ be a dataset of observations in $\mathbb{R}^p$ and $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ be the probability density function of a p-variate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We consider the following robust constrained clustering problem for a fixed trimming level α: search for a partition $R_0, R_1, \ldots, R_k$ of indeces $\{1, \ldots, n\}$ with $\#R_0 = \lceil n\alpha \rceil$, centers $\boldsymbol{m}_1, ..., \boldsymbol{m}_k$ in $\mathbb{R}^p$, symmetric positive semidefinite $p \times p$ scatter matrices $\boldsymbol{S}_1, ..., \boldsymbol{S}_k$ and weights $p_1, ..., p_k$ with $p_j \in [0,1]$ and $\sum_{j=1}^{k} p_j = 1$, which maximizes

$$\sum_{j=1}^{k} \sum_{i \in \mathbb{R}_j} \log\left(p_j \phi(\boldsymbol{x}_i; \boldsymbol{m}_j, \boldsymbol{S}_j)\right). \tag{1}$$

The direct maximization of (1) without any constraint on the scatter matrices is not a well-defined problem. To address this issue, an eigenvalue ratio constraint on $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_k$ is introduced:

$$\frac{\max_{j,l} \lambda_l\left(\boldsymbol{S}_j\right)}{\min_{j,l} \lambda_l\left(\boldsymbol{S}_j\right)} \leq c. \tag{2}$$

Here, $\lambda_l\left(\boldsymbol{S}_j\right)$ for $l = 1, \ldots, p$ represents the set of eigenvalues of the scatter matrix $\boldsymbol{S}_j$, $j = 1, \ldots, k$, and $c \geq 1$ is a constant controlling the strength of the constraint (2), where the smaller the value of c is, the stronger the restriction imposed on the solution.

The maximization (1) under the eigenvalue constraint (2) leads to the TCLUST methodology [3], a fast and efficient algorithm for robust constrained clustering, which shows good performance on low-dimensional data.

The algorithm proceeds through several steps, including initialization, concentration step and target function evaluation. In the initialization, the algorithm is started multiple times with different random assignments and the initial cluster parameters are computed. Then, the following steps are executed until convergence:

- **E-step:**
  For each observation $\boldsymbol{x}_i$, the quantities

  $$D_j\left(\boldsymbol{x}_i; \theta\right) = p_j \phi\left(\boldsymbol{x}_i; \boldsymbol{m}_j, \boldsymbol{S}_j\right)$$

  for $j = 1, \ldots, k$, are computed, with $\theta = (p_1, \ldots, p_k, \boldsymbol{m}_1, \ldots, \boldsymbol{m}_k, \boldsymbol{S}_1, \ldots, \boldsymbol{S}_k)$ as the set of cluster parameters in the current iteration of the algorithm.
- **C-step:**
  The $\lceil n\alpha \rceil$ observations $\boldsymbol{x}_i$ with the smallest values of

  $$\begin{aligned} &D\left(\boldsymbol{x}_i; \theta\right) = \\ &= \max\left\{D_1\left(\boldsymbol{x}_i; \theta\right), \ldots, D_k\left(\boldsymbol{x}_i; \theta\right)\right\} \end{aligned} \tag{3}$$

  are discarded as possible outliers (for this iteration). Each remaining observation $\boldsymbol{x}_i$ is then assigned to a cluster j such that $D_j\left(\boldsymbol{x}_i; \theta\right) = D\left(\boldsymbol{x}_i; \theta\right)$. This yields a partition $R_0, R_1, \ldots, R_k$ of $\{1, \ldots, n\}$ holding the indexes of the trimmed observations in R_0, and the indexes of the observations belonging to cluster j in R_j, for $j = 1, \ldots, k$.
- **M-step:**
  The parameters are updated, based on the non-discarded observations and their cluster assignments. At this point, it is crucial to properly enforce the constraints on the cluster scatter matrices.

Finally, the target function is evaluated to select the best parameters.

2. Well-conditioned estimators for high-dimensional covariance matrices

In this chapter we will introduce a series of regularized covariance matrix estimators to address the challenge of estimating high-dimensional covariance matrices within the TCLUST methodology.

2.1. Minimum Regularized Covariance Determinant estimator

The Minimum Regularized Covariance Determinant (MRCD) approach searches for an h-subset of the data whose regularized covariance matrix has the lowest possible determinant [2].

The variables are standardized before proceeding. The regularized covariance matrix, denoted as $\mathbf{K}(H)$, is a key element in the MRCD approach. It incorporates two essential components: a predetermined target matrix $\mathbf{T}$, which is well-conditioned, symmetric and positive definite, and a scalar weight coefficient ρ. This regularization parameter, ranging from 0 to 1, plays a crucial role in controlling the influence of the target matrix on the result. In particular, $\mathbf{K}(H)$ is defined as follows:

$$\mathbf{K}(H) = \rho \mathbf{T} + (1 - \rho) c_\alpha \boldsymbol{S}_U(H), \tag{4}$$

where $\boldsymbol{S}_U(H)$ is the sample covariance matrix for the the standardized data, and c_α is a consistency factor that depends on the trimming percentage $\alpha = (n - h)/n$.

The MRCD subset H_{MRCD} is defined by minimizing the determinant of the regularized covariance matrix $\mathbf{K}(H)$ in (4):

$$H_{MRCD} = \underset{H \in \mathcal{H}_h}{\operatorname{argmin}} \left(\det(\mathbf{K}(H))^{1/p}\right).$$

Once H_{MRCD} is determined, the MRCD location and scatter estimates of the original data matrix are computed as

$$\begin{aligned} \boldsymbol{m}_{MRCD} &= \boldsymbol{\nu}_X + \boldsymbol{D}_X \boldsymbol{m}_U(H_{MRCD}) \\ \mathbf{K}_{MRCD} &= \boldsymbol{D}_X \mathbf{Q} \mathbf{\Lambda}^{1/2} [\rho \boldsymbol{I} + (1-\rho) c_\alpha \\ &\qquad \boldsymbol{S}_W(H_{MRCD})] \mathbf{\Lambda}^{1/2} \mathbf{Q}^T \boldsymbol{D}_X. \end{aligned}$$

2.2. Linear Shrinkage estimator of Ledoit-Wolf

The goal of the Ledoit-Wolf linear shrinkage estimator [4] is to find the well-conditioned estimator for $\mathbf{\Sigma}$ as the linear combination $\mathbf{\Sigma}^* = \rho_1 \boldsymbol{I} + \rho_2 \boldsymbol{S}$ that minimizes the expected quadratic loss $E\left[\|\mathbf{\Sigma}^* - \mathbf{\Sigma}\|^2\right]$. We introduce four scalar functions of $\mathbf{\Sigma}$: μ, α^2, β^2 and δ^2, which are crucial for finding the optimal linear combination $\mathbf{\Sigma}^*$. A key theorem is presented, demonstrating that:

$$\mathbf{\Sigma}^* = \frac{\beta^2}{\delta^2} \mu \boldsymbol{I} + \frac{\alpha^2}{\delta^2} \boldsymbol{S}, \qquad E\left[\|\mathbf{\Sigma}^* - \mathbf{\Sigma}\|^2\right] = \frac{\alpha^2 \beta^2}{\delta^2}.$$

Here, we are considering the Frobenius norm $\|\boldsymbol{A}\| = \sqrt{\operatorname{tr}\left(\boldsymbol{A}\boldsymbol{A}^T\right)/p}$. However, this estimator relies on knowledge of the four scalar functions, which are typically unknown in practice. To address this limitation, the section introduces a consistent estimation framework based on general asymptotics. It explores a sequence of statistical models, considering the spectral decomposition of covariance matrices. The consistent estimators for the four scalar functions are derived, allowing for the calculation of an efficient unbiased estimator, denoted as $\boldsymbol{S}_n^*$. The most important result is the following: the efficient unbiased estimator $\boldsymbol{S}_n^*$ has uniformly minimum quadratic risk asymptotically among all the linear combinations of the identity with the sample covariance matrix, including those that are efficient unbiased estimators, and even those that use hindsight knowledge of the true covariance matrix. Thus, it is legitimate to say that $\boldsymbol{S}_n^*$ is an asymptotically optimal linear shrinkage estimator of the covariance matrix $\mathbf{\Sigma}$ with respect to quadratic loss under general asymptotics.

2.3. Sparse CovGlasso estimator

Suppose observations come from a p-variate Gaussian distribution with zero mean and covariance matrix $\mathbf{\Sigma}$. The log-likelihood is

$$\ell(\mathbf{\Sigma}) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log(\det(\mathbf{\Sigma})) - \frac{n}{2}\operatorname{tr}(\mathbf{\Sigma}^{-1}\boldsymbol{S})$$

and the goal of the sparse CovGlasso estimator is to find $\mathbf{\Sigma}$ positive definite that minimizes minus the penalized log-likelihood:

$$\log(\det(\mathbf{\Sigma})) + \operatorname{tr}(\mathbf{\Sigma}^{-1}\boldsymbol{S}) + \lambda\|\boldsymbol{P} * \mathbf{\Sigma}\|_1, \qquad (5)$$

where we define $\|\boldsymbol{A}\|_1 = \sum_{ij}|A_{ij}|$. We are adding the lasso penalty of the form $\lambda\|\boldsymbol{P} * \mathbf{\Sigma}\|_1$ to the likelihood, where λ is the lasso regularization parameter, $\boldsymbol{P}$ is an arbitrary matrix with non-negative elements, and $*$ denotes the elementwise multiplication.
Initially, Bien and Tibshirani [1] proposed a majorize-minimize approach to approximately minimize (5), but two years later Wang [5] suggested a new algorithm, showing it has several advantages with respect to the previous one, including simplicity, computational speed and numerical stability. This section presents the minimization of the objective function (5) using the coordinate descent algorithm introduced by Wang.

3. Work development

The initial proposal aims to enhance the TCLUST methodology by integrating the Minimum Regularized Covariance Determinant estimator, addressing the challenge of robust clustering for high-dimensional data. TCLUST employs sample covariance matrices with eigenvalue ratio constraints as robust covariance matrix estimators. However, this approach becomes inadequate in high dimensional settings where the number of variables may exceed the available observations. This can result in unreliable covariance estimates, sensitivity to data noise, highly correlated variables, and numerical instability. To mitigate these issues, the MRCD estimator may replace the sample covariance estimator, providing increased stability and robustness against singularity.
The proposal involves developing a new algorithm based on the TCLUST, and addresses its limitations in high-dimensional data handling. This analysis involves a deep dive into the subfunctions of the TCLUST algorithm, focusing on improving initialization procedure, cluster assignments and parameter estimation. In particular, the initialization procedure is updated to address high-dimensional challenges by randomly assigning observations to clusters (designating some observations as outliers) and com-

puting initial cluster parameters. A key modification is the use of MRCD for covariance matrix estimation, enhancing its stability and reliability. Additionally, a threshold is introduced to guarantee that clusters maintain a minimum size, preventing the formation of empty or excessively small clusters.

In the main function, we carry out several initializations, all of them involving multiple iterations, where the following actions are repeated by calling their corresponding subfunctions:

- Assignment of data points to clusters by computing likelihoods for observations in clusters and updating assignments based on these likelihoods. Each observation is assigned to the cluster that maximizes its likelihood, and the $\lceil n\alpha \rceil$ observations with the smallest values of (3) are discarded as outliers.
- Computation of the value for objective function (1).
- Estimation of mean vectors and covariance matrices for each cluster using the MRCD approach.

At the conclusion of this nested loop, we select the best initialization, which is the one resulting in the highest objective function value. The final result is an object that includes cluster centers, covariance matrices, assignments and other relevant information.

The next steps involve creating a new objective function that incorporates MRCD objective functions for each cluster, potentially reformulating the MRCD problem in terms of likelihood. This step aims to make the proposed method likelihood-based.

We begin by distinguishing between likelihood-based and heuristic methodologies. Likelihood-based methodologies are rooted in probability and statistics, relying on formal probabilistic models. They involve clear assumptions about data distributions and aim to estimate model parameters that maximize the likelihood of observed data. These methods provide interpretable results with statistical inferences. Heuristic methodologies, on the other hand, use practical rules and strategies to solve complex problems, often without formal models or precise distributions. They prioritize finding satisfactory solutions efficiently, making them suitable for large-scale problems.

In this context, we investigate whether the MRCD estimation problem can be expressed in terms of likelihood. We use the likelihood of a multivariate Gaussian distribution and establish the maximum likelihood estimators for mean and covariance matrix. However, we introduce a term to incorporate regularization, resulting in the regularized version of $\mathbf{\Sigma}$ as $\mathbf{K} = \rho\mathbf{T} + (1-\rho)c_\alpha\mathbf{\Sigma}$. Consequently, the problem becomes to minimize:

$$\sum_{i=1}^{n} w_i \left(\ln|\mathbf{K}| + d\ln(2\pi) + \mathrm{MD}^2\left(\boldsymbol{x}_i; \boldsymbol{\mu}, \mathbf{K}\right)\right) \quad \text{under the constraint that} \sum_{i=1}^{n} w_i = h. \tag{6}$$

Since the second term of (6) is a constant, to prove that minimizing (6) is equivalent to minimizing the determinant of

$$\widehat{\mathbf{K}}_{MLE} = \rho\mathbf{T} + (1-\rho)c_\alpha\widehat{\mathbf{\Sigma}}_{MLE} =$$

$$= \rho\mathbf{T} + (1-\rho)c_\alpha\left(\frac{1}{h}\sum_{i=1}^{n} w_i \left(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{MLE}\right)\left(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{MLE}\right)^T\right),$$

we need to prove that the third term of (6) is also a constant. Our analysis concludes that it is not a constant, thus the MRCD estimation problem cannot be reformulated in terms of likelihood and our MRCD in TCLUST methodology remains heuristic in nature.

As next step, we develop another algorithm based on the TCLUST framework, incorporating a Gaussian-based covariance matrix estimator called the CovGlasso estimator. This approach aims to create a likelihood-based methodology for robust clustering in high-dimensional settings.

The primary objective is to construct an objective function that reflects the likelihood-based nature of the methodology. We can formulate this objective function as the summation of k minus penalized log-likelihoods, each representing the individual objective function (5) of the CovGlasso methodology for each cluster:

$$\sum_{j=1}^{k} \left(\log\left(\det\left(\widehat{\mathbf{\Sigma}}_j\right)\right) + \mathrm{tr}\left(\widehat{\mathbf{\Sigma}}_j^{-1}\boldsymbol{S}_j\right) + \lambda\left\|\boldsymbol{P} * \widehat{\mathbf{\Sigma}}_j\right\|_1\right).$$

The objective function of our new methodology needs to be collectively minimized, as it is the sum of k CovGlasso objective functions, each requiring minimization.
The initialization process is crucial. A new type of initial cluster assignments, involving the application of TCLUST on a subset of the original variables, is introduced. This approach offers improved initialization precision and computational efficiency compared to random assignment.
Parameter estimation within clusters is enhanced by incorporating the CovGlasso estimator. The lasso regularization ensures stability and avoids degenerate solutions. In this regard, two additional input parameters are introduced: λ, the lasso regularization parameter, and $\boldsymbol{P}$, the lasso regularization matrix. It is important to note that, in an iterative setting, λ remains constant throughout the algorithm execution, ensuring the consistency of the likelihood-based approach.
In the previously developed heuristic methodology, where the MRCD estimator has been incorporated into the TCLUST algorithm, a single issue arises: this approach presents a doubly-robust extension. Outliers are effectively addressed not only within the clustering procedure itself, but also through the modified M-step that makes use of MRCD to compute the regularized covariance matrices, exploiting its robust-based estimation. Therefore, our intention is to replace the MRCD estimator with the linear shrinkage estimator proposed by Ledoit and Wolf, which can be seen as a particular case of the former, where the subset H in equation (4) corresponds to the entire sample, the target matrix $\mathbf{T}$ is the identity matrix, and the data does not require initial standardization. The Ledoit-Wolf estimator lacks robustness to outliers, as it incorporates them in its computation. Nonetheless, considering its incorporation into the TCLUST algorithm, it becomes logical to use it, given that the robustness is already enforced by the TCLUST procedure. We therefore replace the original MRCD estimator within the M-step with the Ledoit-Wolf approach for covariance matrix estimation, keeping all the remaining steps unchanged, except for the initialization procedure, which is improved as done for CovGlasso in TCLUST.

We finally end up with two distinct algorithms for robust clustering in high-dimensional settings: a heuristic methodology called Ledoit-Wolf in TCLUST, and a likelihood-based methodology, the CovGlasso in TCLUST. These methodologies will be tested and compared using simulated and real-world data in the next chapter.

4. Data analysis

We rigorously assess and compare the two methodologies we have developed for robust clustering in high-dimensional data scenarios. Our evaluation encompasses both simulated and real-world data, with a particular focus on the handwritten digits recognition problem. Our goal is to validate the efficacy of our algorithms in accurately clustering digits while effectively identifying outliers.

4.1. Simulated data

To evaluate our algorithms, we begin with simulated data. We generate synthetic data, simulating a three-component mixture distribution, with each component modeled as a Gaussian distribution. Additionally, we introduce outliers by using a separate Gaussian component. For each Gaussian distribution, we employ distinct mean vectors while sharing a common covariance matrix. This approach allows for a clear visualization of our simulated data in the first two dimensions of the feature space. Indeed, it results in the generated data points from the distinct distributions being closely situated and distinctly separated from each other, as shown in the scatter plot presented in Figure 1.

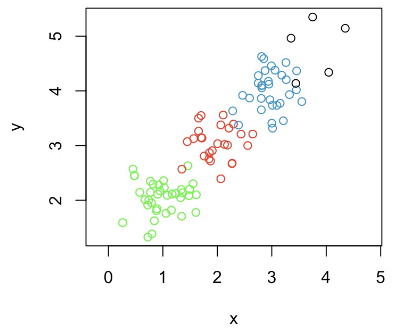

Figure 1: Simulated data in the first two dimensions of the feature space.

Now, we apply our two robust clustering algorithms to this simulated dataset with specifically chosen input parameter values. The results are listed below:

- Both algorithms achieve an exceptional 100% accuracy in detecting outliers and assigning data points to their respective clusters. This outstanding performance demonstrates their robustness.
- The precise cluster assignment indicates the ability of our metodologies to identify patterns within the data, even in presence of noise and outlying units.
- Visual representations in Figure 2 show the same estimated cluster assignments with slightly different elliptical cluster shapes, due to the distinct covariance matrix estimation techniques used in each algorithm.

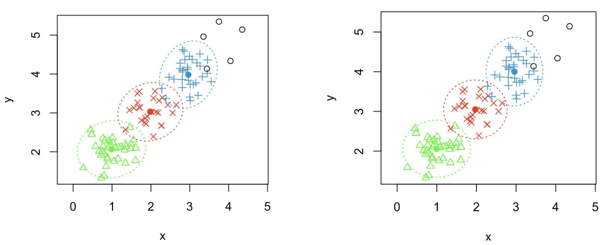

Figure 2: Clustering results using Ledoit-Wolf in TCLUST (on the left) and CovGlasso in TCLUST (on the right).

The evaluation on simulated data confirms the outstanding performance of our robust clustering algorithms, underscoring their robustness and effectiveness when dealing with high-dimensional simulated data. Now, we transition to real-world data, specifically addressing the challenge of handwritten digit recognition, where we anticipate greater complexity.

4.2. Real-world data: the handwritten digits recognition problem

This section delves into the practical application of our robust clustering algorithms to real-world high-dimensional data. We concentrate on two datasets designed for handwritten digit recognition, arising from the USPS dataset available through the UCI Machine Learning Repository. It contains images of handwritten digits from 0 to 9, partitioned into a 16×16 grid, resulting in 256 pixels as features, as shown in Figure 3.

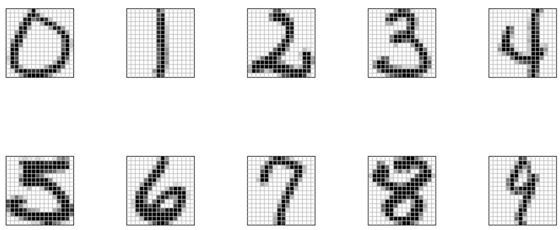

Figure 3: Visual representation of digits 0 to 9 from the handwritten digits dataset.

We start our analysis of the USPS dataset by plotting the multivariate means for all the digits. This initial step aims to identify two distinct groups: one comprising quite distinguishable digits and the other composed of digits that exhibit the highest similarities among themselves. This allows us to apply our algorithms to two different subsets of the original USPS dataset, each representing a different level of complexity. In the more complex one, in addition to the challenges of high-dimensional data and presence of outliers, there arises also the issue of limited separation between classes.

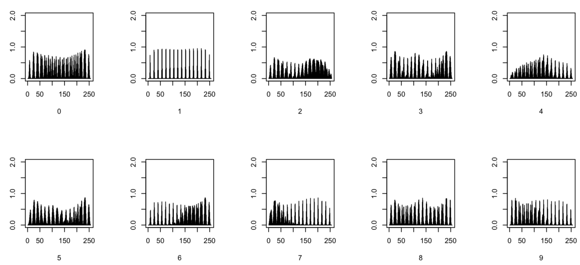

Figure 4: Multivariate means for all digits within the USPS dataset.

It is evident from Figure 4 that digits 0, 1 and 4 exhibit clearly distinct behaviors, each diverging significantly from the others, while digits 3, 5 and 8 display remarkably similar multivariate means.
Our approach involves the application of our two robust clustering methodologies, Ledoit-Wolf in TCLUST and CovGlasso in TCLUST, with the primary goal of assessing their effectiveness and robustness. In the case of more easily distinguishable digits (0, 1 and 4), both algorithms exhibit remarkable clustering results, with Ledoit-Wolf in TCLUST yielding an overall accuracy of 90.3% and ARI of 0.729, and CovGlasso in TCLUST boasting an overall accuracy of 96.8%

and ARI of 0.905. In particular, both methodologies excel in identifying the five anomalous units present in the processed dataset, accurately labeling them as outliers, as shown in Table 1.

group	0	1	4	out
0	45	3	2	0
1	0	50	0	0
4	1	9	40	0
out	0	0	0	5

(a) Ledoit-Wolf in TCLUST

group	0	1	4	out
0	48	0	2	0
1	0	49	1	0
4	0	2	48	0
out	0	0	0	5

(b) CovGlasso in TCLUST

Table 1: Contingency tables for comparisons between estimated cluster labels and true labels (0, 1 and 4).

For the more challenging digits (3, 5 and 8), CovGlasso in TCLUST outperforms Ledoit-Wolf in TCLUST, achieving an overall accuracy of 69.7% and ARI of 0.385.

group	3	5	8	out
3	48	2	0	0
5	29	19	2	0
8	4	10	36	0
out	0	0	0	5

Table 2: Contingency table for comparisons between estimated cluster labels and true labels using CovGlasso in TCLUST (3, 5 and 8).

While the algorithm struggles to correctly identify the true 5s, primarily assigning them to the estimated cluster of 3s as shown in Table 2, the other digits are mostly recognized accurately, and the outliers continue to be detected as well. To conclude, we validate CovGlasso in TCLUST, the methodology demonstrating superior performance in clustering digits 0, 1 and 4, as well as digits 3, 5 and 8, by confirming the decreasing trend in the objective function and the sparsity patterns in the covariance matrices, aligning with our methodology expectations.

5. Conclusions

This thesis presents a solution to the complex problem of robust clustering in high-dimensional data scenarios. We implemented two methodologies: one heuristic, incorporating the Ledoit-Wolf linear shrinkage estimator into TCLUST, and another likelihood-based, employing the sparse CovGlasso estimator within TCLUST. We then applied these methodologies to both simulated and real-world data, to test and evaluate them. In the case of the simulated dataset, both of our algorithms demonstrated exceptional performance by correctly assigning all data points to their respective clusters and identifying all anomalous units within the data. For the dataset of digits 0, 1 and 4, both algorithms performed very well, achieving high overall accuracy and ARI scores. However, on the dataset containing digits 3, 5 and 8, CovGlasso in TCLUST outperformed Ledoit-Wolf in TCLUST, producing satisfactory results despite the inherent challenges of the task, which includes high data dimensionality, the presence of outliers, and significant similarity between classes. In conclusion, CovGlasso in TCLUST emerges as a robust solution for addressing clustering challenges on contaminated high-dimensional data.

References

[1] J. Bien and R. J. Tibshirani. Sparse estimation of a covariance matrix. *Biometrika*, page 807–820, 2011.

[2] K. Boudt, P. J. Rousseeuw, S. Vanduffel, and T. Verdonck. The minimum regularized covariance determinant estimator. *Statistics and Computing*, 30:113–128, 2020.

[3] H. Fritz, L. Garcìa-Escudero, and A. Mayo-Iscar. A fast algorithm for robust constrained clustering. *Computational Statistics and Data Analysis*, 61:124–136, 11 2012.

[4] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, 2 2004.

[5] H. Wang. Coordinate descent algorithm for covariance graphical lasso. *Statistics and Computing*, 24:521–529, 2013.