ROBUST DETECTION FOR CLUSTER ANALYSIS

Michael Fau β^* , *Michael Muma*^{*}, *Freweyni K. Teklehaymanot*^{*,†}, and Abdelhak M. Zoubir^{*,†}

*Signal Processing Group, Technische Universität Darmstadt, Germany †Graduate School CE, Technische Universität Darmstadt, Germany

ABSTRACT

The problem of deciding whether a given set of data points forms one cluster or two clusters is investigated from a robust hypothesis testing perspective. It is assumed that a clustering algorithm exists that for both cases calculates cluster assignments and estimates of the corresponding probability density functions. Based on the latter, a statistical hypothesis test for the true number of clusters is formulated. In order to take falsely labeled data points into account, the clusters are then modeled as being contaminated with outliers. This leads to an uncertainty model for the cluster densities of the ε -contamination type, whose corresponding minimax optimal robust detector is well-known and can be implemented using least favorable densities. The performance of this detector under cluster overlap, cluster imbalance, and for different contamination ratios is evaluated numerically and is compared to that of a Bayesian cluster enumeration criterion. Significant performance improvements are shown in all cases.

Index Terms— Clustering, Cluster Enumeration, Cluster Analysis, Rubust Detection

1. INTRODUCTION

A major challenge in cluster analysis is that the number of data clusters is usually unknown. The estimation of the number of clusters, also called cluster enumeration, has been intensively researched [1–17]. Generally speaking, enumeration approaches may be loosely grouped into two categories. The first category constructs cluster validity indices, such as [6–9] which estimate the number of clusters by measuring the compactness and separation of the clusters. The second category uses criterion functions under a probabilistic mixture-model framework, such as [5,11,15] that trade off model complexity and data fit.

Typical challenges for cluster enumeration include cluster overlap and cluster imbalance [2, 5]. For example, a cluster enumeration criterion may not provide a reliable decision for any of the candidate models, which can lead to over-segmentation or undersegmentation. For such scenarios, we propose a robust detector that can be used for a subsequent cluster verification and refinement. More precisely, we design a minimax optimal robust hypothesis test under ε -contamination to decide whether the data forms one or two clusters.

The reasons for focusing on the two cluster case are twofold. First, the two cluster case can be considered an atomic building block, based on which more complex and powerful, e.g., hierarchical algorithms can be designed. Second, limiting the analysis to two clusters allows for building upon existing results in minimax robust detection and forgo complications that arise with robust hypothesis tests for multiple hypotheses. From a practical viewpoint, the proposed robust detector can be used as a tool to improve and verify the results of a given clustering method. For example, cluster overlap can lead to falsely labelled data points in the overlapping region between two clusters. Being designed explicitly for this kind of uncertainty, the robust detector is able to deal with mislabeled data points and shows promising performance, even for severe overlap. On the other hand, in the presence of cluster imbalance, the proposed detector may be used to decide if a small partition in the data is actually a separate cluster, or if it should be fused with a nearby larger cluster.

The proposed robust detector for cluster analysis is detailed in Section 2. The use-cases of cluster overlap and cluster imbalance are investigated via numerical experiments and a real-world example in Section 3. Possible extensions are discussed in Section 4, which also concludes the paper.

2. PROPOSED ROBUST DETECTOR FOR CLUSTER ANALYSIS

Let $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \in \Omega^N$, with N > 1, be a set of observations which contains K^0 clusters. We assume that a clustering algorithm is given that takes the set \mathcal{X} and a positive integer K as input arguments and outputs

- *K* independent and non-empty clusters X_{1:K},..., X_{K:K} that satisfy ⋃_{k=1}^K X_{k:K} = X and X_{k:K} ∩ X_{j:K} = Ø for all k ≠ j.
- K estimates p̂_{1:K},..., p̂_{K:K} of the density functions that are assumed to have generated the observations in the respective clusters.

Based on the estimated densities $\hat{p}_{k:K}$, statistical hypotheses about the true number of clusters can be formulated¹ as

$$\mathcal{H}_K: \quad p(\mathcal{X}) = \prod_{k=1}^K \hat{p}_{k:K}(\mathcal{X}_{k:K}), \tag{1}$$

where K ranges from one to a maximum number of clusters that is determined by the test designer. We assume that the true number of clusters K^0 lies within the range of considered clusters. For the special case considered in this paper, namely testing whether the data forms one or two clusters, the hypotheses in (1) become

$$\begin{aligned} \mathcal{H}_{1} : & p(\mathcal{X}) = \hat{p}_{1:1}(\mathcal{X}_{1:1}), \\ \mathcal{H}_{2} : & p(\mathcal{X}) = \hat{p}_{1:2}(\mathcal{X}_{1:2})\hat{p}_{2:2}(\mathcal{X}_{2:2}). \end{aligned}$$
(2)

The work of M. Muma is supported by the 'Athene Young Investigator Programme' of Technische Universität Darmstadt. The work of F. K. Teklehaymanot is supported by the 'Excellence Initiative' of the German Federal and State Governments and the Graduate School of Computational Engineering at Technische Universität Darmstadt.

¹Formulating the hypotheses in terms of estimated/empirical distributions is a shorthand notation that is used for compactness. More accurately, hypotheses are formulated in terms of the true distributions, which are unknown, so that they need to be estimated from the data in an additional step.

Since the hypotheses in (2) are simple, the optimal test for \mathcal{H}_1 against \mathcal{H}_2 , in the sense of Neyman and Pearson, is a likelihood ratio test of the form

$$\delta_{2}(\mathcal{X}) = \begin{cases} 0, \quad l(\mathcal{X}) < \lambda \\ \kappa, \quad l(\mathcal{X}) = \lambda \\ 1 \quad l(\mathcal{X}) > \lambda \end{cases}$$
(3)

where $\delta_2 \colon \Omega^N \to [0, 1]$ denotes the probability of accepting \mathcal{H}_2 , given the observations $\mathcal{X}, l \colon \Omega^N \to \mathbb{R}$ denotes the log-likelihood ratio

$$l(\mathcal{X}) = \log \frac{\hat{p}_{1:2}(\mathcal{X}_{1:2})\hat{p}_{2:2}(\mathcal{X}_{2:2})}{\hat{p}_{1:1}(\mathcal{X}_{1:1})},$$
(4)

 $\lambda \in \mathbb{R}$ denotes the decision threshold, and the randomization $\kappa \in [0, 1]$ can be chosen arbitrarily.

The problem that arises when implementing the above test in a straightforward manner is the following: while the estimate for $\hat{p}_{1:1}$ is based on a trivial, unambiguous cluster assignment, the estimates $\hat{p}_{1:2}$ and $\hat{p}_{2:2}$ are based on potentially erroneous cluster assignments. That is, $\hat{p}_{1:2}$ and $\hat{p}_{2:2}$ are subject to more uncertainty than $\hat{p}_{1:1}$. Procedures that do not take this uncertainty imbalance into account are bound to overestimate the likelihood of \mathcal{H}_1 , i.e., they are biased towards the single cluster hypothesis, especially when the clusters overlap. This effect can be seen in the Bayesian information criterion (BIC) in [5] as well as in the non-robust test introduced above and is illustrated with simulated and real-data examples in Section 3.

Naturally, there are various options to take the additional uncertainty under \mathcal{H}_2 into account. For example, one could robustify the estimators used to obtain $\hat{p}_{1:2}$ and $\hat{p}_{2:2}$ so as to reduce the influence of erroneously assigned observations on the estimated distributions [18, 19]. However, the estimators are often tightly interlinked with the clustering algorithm so that one cannot be changed without affecting the other. The approach proposed here avoids this problem by not trying to robustify the estimators, but the subsequent test. More precisely, from a detection point of view, incorrectly assigned data points can be interpreted as observations that were drawn from a different distribution. That is, observations that were erroneously assigned to $\mathcal{X}_{1:2}$ are not drawn from $\hat{P}_{1:2}$, but from $\hat{P}_{2:2}$ and vice versa. In robust statistics, this type of corruption in a data set is known as ε -contamination, where ε is referred to as contamination ratio. It defines the fraction of observations that are assumed to be outliers, which here corresponds to the fraction of misassigned observations in the clusters.

A minimax optimal hypothesis test under ε -contamination is designed by replacing the nominal densities, in this case $\hat{p}_{1:2}$ and $\hat{p}_{2:2}$, with the *least favorable densities*, here denoted by $\hat{q}_{1:2}$ and $\hat{q}_{2:2}$. In a nutshell, the latter are obtained by maximizing the error probabilities with respect to the distributions from which the outliers are assumed to be drawn. The robust test is then designed such that it achieves the best possible error probabilities under the worst possible outlier distributions; see [20, 21] for details.

The detector proposed in this paper is based on the same design principle, but with an important difference. Since the cluster under \mathcal{H}_1 does not contain misassigned observations, it is not affected by outliers. Consequently, the least favorable distributions are not those that maximize the probability of confusing \mathcal{H}_1 and \mathcal{H}_2 , but those that maximize the probability of confusing the two clusters under \mathcal{H}_2 .

Following the arguments detailed in [21], the least favorable densities can be shown to be of the form

$$\hat{q}_{1:2}(x) = \max\{c_1\hat{p}_{2:2}(x), (1-\varepsilon_1)\hat{p}_{1:2}(x)\},\ \hat{q}_{2:2}(x) = \max\{c_2\hat{p}_{1:2}(x), (1-\varepsilon_2)\hat{p}_{2:2}(x)\},\ (5)$$

where the maximum is taken element-wise, $\varepsilon_1, \varepsilon_2$ denote the contamination ratios, and c_1, c_2 need to be chosen such that $\hat{q}_{1:2}$ and $\hat{q}_{2:2}$ are valid densities, i.e., they integrate to one. The corresponding robust test for \mathcal{H}_1 against \mathcal{H}_2 is then given by the decision rule in (3), where the nominal log-likelihood ratio in (4) is replaced by the robust log-likelihood ratio

$$l(\mathcal{X}) = \log \frac{\hat{q}_{1:2}(\mathcal{X}_{1:2})\hat{q}_{2:2}(\mathcal{X}_{2:2})}{\hat{p}_{1:1}(\mathcal{X}_{1:1})}.$$
(6)

In the next section, the reliability of this detector is evaluated numerically under different assumptions and scenarios.

3. NUMERICAL RESULTS

In this section, several examples are shown which provide a proofof-concept that robust detection can be a useful tool in cluster verification and enumeration. The scenarios are deliberately kept simple in order to allow isolating the effects of different parameters on the detection performance. A more rigorous performance analysis is beyond the scope of this paper and will be the subject of a future publication.

The proposed detector needs to be paired with a clustering algorithm that provides estimates of the clusters and their associated densities. Here, the clustering method detailed in [5] is used, which is a state-of-the-art approach based on expectation maximization (EM) [22]. Moreover, the BIC derived in [5] provides a reference method against which the proposed detector is compared. The BIC [5] formulates the problem of estimating the number of clusters in a data set as the maximization of the posterior probability of the candidate models.

All results presented in this section were obtained via Monte Carlo simulations using 1000 runs. The simulation setup is as follows: under \mathcal{H}_2 , the observations are generated by independently drawing $N_1 = |\mathcal{X}_{1:2}|$ samples from a bivariate ($\Omega = \mathbb{R}^2$) Gaussian distribution with mean $\mu_{1:2}$ and covariance $\Sigma_{1:2}$ and drawing $N_2 = |\mathcal{X}_{2:2}|$ samples from a Gaussian distribution with mean $\mu_{2:2}$ and covariance $\Sigma_{2:2}$. Under \mathcal{H}_1 , the observations are generated by drawing $N = N_1 + N_2$ samples from a single Gaussian distribution with mean

$$\mu_{1:1} = \alpha_1 \mu_{1:2} + \alpha_2 \mu_{1:2}$$
(7)

and covariance

$$\Sigma_{1:1} = \Phi_{1:1} - \mu_{1:1} \mu_{1:1}^{1}, \qquad (8)$$

where $\alpha_1 = N_1/N$, $\alpha_2 = N_2/N$, and

$$\boldsymbol{\Phi}_{1:1} = \alpha_1 \left(\boldsymbol{\Sigma}_{1:2} + \boldsymbol{\mu}_{1:2} \boldsymbol{\mu}_{1:2}^{\mathrm{T}} \right) + \alpha_2 \left(\boldsymbol{\Sigma}_{2:2} + \boldsymbol{\mu}_{2:2} \boldsymbol{\mu}_{2:2}^{\mathrm{T}} \right). \quad (9)$$

That is, the mean and covariance of $\hat{p}_{1:1}$ are chosen such that they match the mean and covariance of $\alpha_1 p_{1:2} + \alpha_2 p_{2:2}$. Note that the bivariate Gaussian distribution is used merely for simplicity; both the clustering algorithm and the proposed detector extend to the multivariate case and to non-Gaussian distributions.

Throughout the section, the empirical probability of correctly detecting the presence of a second cluster is referred to as *detection rate* and the empirical probability of incorrectly detecting a second cluster is referred to as *false alarm rate*.

3.1. Cluster overlap

The purpose of the first experiment is to asses how well overlapping clusters can be identified by the robust detector. To this end, the



Fig. 1. ROCs of the proposed detector for different cluster center distances.

	Cluster Center Distance d						
	3.2	2.8	2.4	2.0	1.6		
False Alarm Rate	1%	0.2%	0.7%	0.7%	0.6%		
Detection Rate	5.9%	0.5%	0.0%	0.1%	0.3%		

 Table 1. False alarm and detection rates of the BIC in [5] for different cluster center distances.

cluster means under \mathcal{H}_2 are parametrized as

$$\boldsymbol{\mu}_{1:2} = -\boldsymbol{\mu}_{2:2} = \begin{bmatrix} d/2 & 0 \end{bmatrix}^{\mathrm{T}}, \tag{10}$$

where d determined the distance between the cluster centers and in turn the cluster overlap. Both clusters are assumed to be white and of unit covariance. That is, $\Sigma_{1:2} = \Sigma_{2:2} = I$, and the number of samples per cluster is set to $N_1 = N_2 = 100$.

In Fig. 1, the receiver operating characteristic (ROC) of the proposed detector with $\varepsilon_1 = \varepsilon_2 = 0.3$ is shown for different degrees of cluster overlap. By inspection, the detection performance is excellent for small overlaps, but deteriorates for large overlaps. In order to get an idea of how large the overlap in the different scenarios really is, two representative examples for observations under \mathcal{H}_1 and \mathcal{H}_2 with d = 2.4 are shown in Fig. 2. As can be seen, the point clouds are highly similar and there is considerable overlap between the two clusters in the lower plot. Nevertheless, for d = 2.4, the proposed detector is able to reliably separate the clusters, with detection rates at around 80% for false alarm rates of 10%.

The BIC, in contrast, shows a strong bias towards the single cluster hypothesis. That is, while it admits very low false alarm rates, it does not reliably detect the presence of a second cluster. This can be seen in Table 1, where false alarm rates and detection rates of the BIC are listed for different degrees of cluster overlap. Even for the case d = 3.2, for which the clusters can be separated almost perfectly by the robust detector, the BIC does not achieve detection rates above 5%. This observation confirms the theoretical considerations in the previous section. However, it should also be highlighted that the clusters in this example are particularly difficult to separate using a Gaussian BIC since Gaussian distributions fit the data well under both hypotheses.



Fig. 2. Representative example of observations under \mathcal{H}_1 (top) and \mathcal{H}_2 (bottom) with cluster centers at $[\pm 1.2 \ 0]^T$.

	Cluster Imbalance N_1/N_2					
	1	2	3	4	5	
False Alarm Rate	0.2%	0.8%	0.9%	0.7%	0.8%	
Detection Rate	0.5%	0.4%	3.3%	6.4%	9.3%	

 Table 2. False alarm and detection rates of the BIC in [5] for different degrees of cluster imbalance.

3.2. Cluster imbalance

Besides the cluster overlap, the cluster imbalance, i.e., the difference in the number of cluster elements, has been shown to have a strong effect on the cluster enumeration [2, 5]. Therefore, in the second experiment, the ratio of observations per cluster, N_1/N_2 , is varied from 1 to 5, while N = 200, d = 2.8, and $\varepsilon_1 = \varepsilon_2 = 0.3$ are held constant. The corresponding ROCs are depicted in Fig. 3. The loss in performance caused by an increasing cluster imbalance is clearly visible. However, for moderate cluster imbalances $(N_1/N_2 \leq 3)$, the two hypotheses can still be separated with a high reliability. Since for d = 2.8 there also is a significant cluster overlap, these results are promising overall, especially compared to the BIC, whose detection and false alarm rates are given in Table 2. Interestingly, the detection rate of the BIC increases with the cluster imbalance. This effect is attributed to the difference in the structure of the clusters when there is cluster imbalance. In such cases the proposed detector might not be superior to the BIC anymore.

3.3. Contamination ratio

So far, the contamination ratios $\varepsilon_1, \varepsilon_2$ have been kept fixed at a value of 0.3, which corresponds to 30% erroneously assigned observations per cluster. Our experiments suggest that, without prior



Fig. 3. ROCs of the proposed detector for different degrees of cluster imbalance.



Fig. 4. Detection rates of the proposed detector vs. contamination ratio for different false alarm rates.

knowledge about the structure of the clusters, this value leads to an overall good performance. However, if prior knowledge is available, it can be used to further improve the detection rates. This is illustrated in Fig. 4, where the detection rate of the proposed detector is plotted vs. the contamination ratio for different false alarm rates and d = 2.4, $N_1 = N_2 = 100$. For all depicted false alarm rates, the beneficial effect of taking the possibility of outliers into account in the test design is clearly visible, with the best robust test achieving detector ($\varepsilon_1 = \varepsilon_2 = 0$). The detection rates reach a maximum between contamination rates of approximately 0.2 and 0.3. On the one hand, this shows that there is indeed a risk of over-robustification. On the other hand, the relatively flat performance profiles show that the choice of the contamination rates is not critical, as long as they stay within an appropriate interval.

3.4. Iris flower data set

The final experiment is to apply the proposed robust detector to a real-world data set, namely the iris flower data set introduced by Fisher [23, 24]. Here only two types of iris flowers and two features are considered, namely the sepal and the petal widths. The two clusters are depicted in the upper plot of Fig. 5, where a moderate cluster overlap can be seen.



Fig. 5. Example of two clusters based on two features from the iris flower data set. The true clusters are shown in the upper plot, the clusters estimated by the EM algorithm [5] in the lower plot.

Applying the proposed detector with $\varepsilon_1 = \varepsilon_2 = 0.3$ to the data set yields a log-likelihood ratio of approximately 40 in favor of the two cluster case. For reasonably chosen thresholds, this corresponds to a correct detection. For comparison, in order to separate the clusters shown in Fig. 2 with a false alarm rate of 5 %, the log-likelihood ratio threshold needs to be set to $\lambda \approx 22$.

As expected, on this reduced set of features, the BIC decides in favor of a single cluster. (Using all available features, the clusters are recognized correctly by the BIC [5].) Interestingly, the underlying EM algorithm does separate the two clusters with a reasonably high accuracy; compare the upper and lower plot in Fig. 5. This discrepancy highlights the need for reliable cluster enumeration methods, even in cases with little cluster overlap or imbalance.

4. CONCLUSIONS AND POSSIBLE EXTENSIONS

The results presented in the previous section provide evidence that robust detection can be a useful tool for cluster verification and enumeration. A natural extension is to consider the case of multiple clusters, which in turn requires the use of robust detectors for multiple hypotheses. This extension is non-trivial and will be investigated in a future publication. In the bigger picture of things, however, we do not see robust detection as a competitor to existing cluster enumeration methods, but rather as a complement. In practice, one needs to be able to efficiently handle high-dimensional data sets and a large numbers of clusters so that an enumeration purely based on robust detection is likely to be too complex to be useful. However, robust detectors might be able to resolve cases where the outcomes of traditional cluster enumeration methods are inconclusive or unreliable. Exploring this symbiosis is another promising area for future research.

5. REFERENCES

- R. Xu and D. Wunsch, "Survey of Clustering Algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.
- [2] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An Extensive Comparative Study of Cluster Validity Indices," *Pattern Recognition*, vol. 46, no. 1, pp. 243–256, 2013.
- [3] G. W. Milligan and M. C. Cooper, "An Examination of Procedures for Determining the Number of Clusters in a Data Set," *Psychometrika*, vol. 50, no. 2, pp. 159–179, 1985.
- [4] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On Clustering Validation Techniques," *Journal of Intelligent Information Systems*, vol. 17, no. 2/3, pp. 107–145, 2001.
- [5] F. K. Teklehaymanot, M. Muma, and A. M. Zoubir, "Bayesian Cluster Enumeration Criterion for Unsupervised Learning," *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5392–5406, 2018.
- [6] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters," *Journal* of Cybernetics, vol. 3, no. 3, pp. 32–57, 1973.
- [7] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, pp. 224–227, 1979.
- [8] T. Caliński and J. Harabasz, "A Dendrite Method for Cluster Analysis," *Communications in Statistics-theory and Methods*, vol. 3, no. 1, pp. 1–27, 1974.
- [9] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.
- [10] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the Number of Clusters in a Dataset via the Gap Statistic," *Journal* of the Royal Statistical Society Series B (Statistical Methodology), vol. 63, pp. 411–423, 2001.
- [11] D. Pelleg and A. Moore, "X-means: Extending K-Means with Efficient Estimation of the Number of Clusters," in *Proceedings of the 17th International Conference on Machine Learning (ICML)*, 2000, pp. 727–734.
- [12] A. Kalogeratos and A. Likas, "Dip-Means: An Incremental Clustering Method for Estimating the Number of Clusters," in Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS), 2012, pp. 2402–2410.
- [13] Y. Feng and G. Hamerly, "PG-Means: Learning the Number of Clusters in Data," in Advances in Neural Information Processing Systems (NIPS) 19, 2007, pp. 393–400.
- [14] G. Hamerly and E. Charles, "Learning the K in K-Means," in Advances in Neural Information Processing Systems (NIPS) 17, 2003, pp. 281–288.
- [15] F. K. Teklehaymanot, M. Muma, J. Liu, and A. M. Zoubir, "In-Network Adaptive Cluster Enumeration for Distributed Classification/Labeling," in *Proceedings of the 24th European Signal Processing Conference (EUSIPCO)*, 2016, pp. 448–452.
- [16] T. Ishioka, "An Expansion of X-Means for Automatically Determining the Optimal Number of Clusters," in *Proceedings of the 4th IASTED International Conference on Computational Intelligence*, 2005, pp. 91–96.

- [17] P. Binder, M. Muma, and A. M. Zoubir, "Gravitational Clustering: A Simple, Robust and Adaptive Approach for Distributed Networks," *Signal Processing*, vol. 149, pp. 36–48, 2018.
- [18] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust Estimation in Signal Processing: A Tutorial-Style Treatment of Fundamental Concepts," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61–80, 2012.
- [19] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*, Cambridge University Press, 2018.
- [20] P. J. Huber, "A Robust Version of the Probability Ratio Test," *The Annals of Mathematical Statistics*, vol. 36, no. 6, pp. 1753– 1758, 1965.
- [21] M. Fauß and A. M. Zoubir, "Old Bands, New Tracks— Revisiting the Band Model for Robust Hypothesis Testing," *IEEE Transactions on Signal Processing*, vol. 64, no. 22, pp. 5875–5886, 2016.
- [22] C. M. Bishop, Pattern Recognition and Machine Learning, Springer Science+Business Media, LLC, New York, USA, 2006.
- [23] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *The Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
- [24] M. Lichman, "UCI Machine Learning Repository," 2013, http://archive.ics.uci.edu/ml.