"Anti–Bayesian" Flat and Hierarchical Clustering Using Symmetric Quantiloids

Anis Yazidi[1], Hugo Lewi Hammer[1], and B. John Oommen[2⋆]

[1] Department of Computer Science, Oslo and Akershus University College of Applied Sciences, Norway
[2] School of Computer Science, Carleton University, Ottawa, Canada

Abstract. A myriad of works has been published for achieving data clustering based on the Bayesian paradigm, where the clustering sometimes resorts to Naïve-Bayes *decisions*. Within the domain of clustering, the Bayesian principle corresponds to assigning the unlabelled samples to the cluster whose mean (or centroid) is the closest. Recently, Oommen and his co-authors have proposed a novel, counter-intuitive and pioneering PR scheme that is radically opposed to the Bayesian principle. The rational for this paradigm, referred to as the "Anti-Bayesian" (AB) paradigm, involves classification based on the non-central *quantiles* of the distributions. The first-reported work to achieve clustering using the AB paradigm was in [1], where we proposed a flat clustering method which assigned unlabelled points to clusters based on the AB paradigm, and where the distances to the respective learned clusters was based on their quantiles rather than the clusters' centroids for uni-dimensional and two-dimensional data. This paper, extends the results of [1] in many directions. Firstly, we generalize our previous AB clustering [1], initially proposed for handling uni-dimensional and two-dimensional spaces, to arbitrary d-dimensional spaces using their so-called *"quantiloids"*. Secondly, we extend the AB paradigm to consider how the clustering can be achieved in hierarchical ways, where we analyze both the Top-Down and the Bottom-Up clustering options. Extensive experimentation demonstrates that our clustering achieves results competitive to the state-of-the-art flat, Top-Down and Bottom-Up clustering approaches, demonstrating the power of the AB paradigm.

1 Introduction

Clustering is the task of grouping data points in a way that elements that exhibit some similarity, or that inherently belong to the same class, end up in the same group. It is a fundamental task in data analysis and inference, and it is, arguably, among the most popular machine learning and data mining techniques [2] [3]. A range of different clustering methods have been proposed and each of them vary with the understanding of what a cluster, actually, is. For instance, density

⋆ *Chancellor's Professor*; *Fellow: IEEE* and *Fellow: IAPR*. This author is also an *Adjunct Professor* with the University of Agder in Grimstad, Norway.

models, such as OPTICS [4] and DBSCAN [5], coalesce most dense regions in the space into a single cluster. As opposed to this, in hierarchical clustering [6] [7], the aim is to arrange the data points into an underlying hierarchy which then determines the various clusters. A third group of clustering algorithms constitute the so-called "centroid" methods where all the points within a computed cluster are represented by a single point, for example the cluster's centroid. The most prominent example of a scheme within this family is the acclaimed k-means clustering algorithm where a centroid is represented by the mean value of the points in the cluster. The central strategy motivating *these* clustering schemes involves classifying *unassigned* data points to the different clusters based on the distances to the means (or centroids) of the clusters. From the above, one can informally see that any specific pattern classification algorithm can be conceptually expanded to yield a clustering scheme. Thus, if we have k previously-determined clusters, an unknown unlabelled sample can be assigned to any one of the k classes by the corresponding classification algorithm, whence the specific cluster can be grown to include this specific sample. Almost all the well-known classifiers involved in pattern classification are based on a Bayesian principle which aims to maximize the *a posteriori* probability. Quite recently, Oommen and his co-authors proposed a completely counter-intuitive paradigm, known as CMQS, the Classification by Moments of Quantile Statistics. CMQS works with a counter-intuitive philosophy, and essentially compares the testing sample with points from each class which are distant from the mean – as opposed to the Bayesian principle which essentially compares it to the clusters' means or the centroids

The question that begged investigation and that was considered open was that of invoking these "Anti–Bayesian" (AB) PR algorithms to design the corresponding clustering algorithms. This is the avenue of research undertaken here. The pioneering steps taken in this direction were reported in [1], where we introduced a novel alternative to the k-means clustering algorithm. The algorithm presented in [1] follows the same steps dictated by a typical k-means clustering algorithm. The main difference, however, is the manner by which it assigns the data points to the already-formed clusters. Indeed, rather than follow a Bayesian classification methodology, it traverses one of the AB-based PR CMQS-based schemes reported earlier. In fact, unlike the k-means clustering strategies that rely on centroid-based criteria, we resort to *quantiles positions distant from the cluster means* [8] [9] [10], which is a strategy just as counter-intuitive and non-obvious as the CMQS schemes themselves. Central to the development of such CMQS clustering algorithms is the concept of a "Quantiloid" We will elaborate on the phenomenon of Quantiloids in the next section. It is pertinent to mention that by working with Quantiloids, we will have effectively extended our previous work [1]. However, apart from doing it in the "vanilla" manner, we shall accomplish it by also invoking hierarchical clustering approaches.

1.1 Structure of the paper

In Section 2, we present the fundamental principles of AB clustering. In Section 3, we demonstrate the development of AB flat clustering in d-dimensional spaces.

The principles of hierarchical AB clustering are given in Section 4. In Section 5, we report our experimental results which compare our AB flat and hierarchical clustering schemes to their Bayesian counterparts. Section 6 concludes the paper.

2 The "Anti–Bayesian" Clustering Solution

2.1 Quantiloids

As alluded to earlier, the solution we propose is based on the concept of "Quantiloids". What then is a Quantiloid? The quantiloid associated with the real number, θ, is, quite simply, for a uni-dimensional distribution, the unique point where the Cumulative Distribution Function (CDF) has the value θ. This is the unique point where the probability mass (i.e., the integral of the Probability Density Function (PDF)) attains the value of θ. While this is an elementary concept for uni-dimensional variables, the concept can be extended for multi-dimensional vectors to be the hyper-surface under which the CDF has the value θ. The goal of this paper is to develop quantiloid-based clustering algorithms that work in an AB paradigm just as the centroid-based clustering algorithms worked within the "Bayesian" paradigm. Indeed, rather than characterizing a cluster by its centroid, we shall attempt to characterize it by its quantiloids, which will then lead to the various AB clustering algorithms.

Although the concept of quantiloids is valid for multi-dimensional vectors, the question of *how* they can be computed and represented is still open. We shall thus restrict ourselves to uni-dimensional quantiloids by processing the multi-dimensional distribution in terms of its uni-dimensional marginals.

2.2 "Anti–Bayesian" Classification Rules

We first summarize the AB classification rules designed and proven in [8], [9] and [10] for uni-dimensional features. To do this, we use the notation that for the j^{th} dimension of the feature vector of class ω_i, $q_p^{i,j}$ is the quantiloid for the value p, i.e., the position where the feature's CDF has a value of p. In the case when both the classes are characterized by only a *single* feature X, q_p^i is ω_i's quantiloid for the value p, i.e. more formally $q_p^i = Pr(X < p|X \in \omega_i)$. Observe that we encounter the cases when the quantiloids overlap (i.e., $q_{1-p}^1 < q_p^2$) or when they do not overlap (i.e., $q_{1-p}^1 > q_p^2$). Using this notation, the uni-dimensional AB classification rules for the testing sample x^* are:

Case 1: When the quantiloids are non-overlapping (see Figure 1 on the left):

$$\begin{array}{ll} \text{If } x^* < q_{1-p}^1 & \Rightarrow x^* \in \omega_1; \\ \text{If } x^* > q_p^2 & \Rightarrow x^* \in \omega_2; \\ \text{If } (q_{1-p}^1 < x^* < q_p^2) \ \wedge (\|x^* - q_{1-p}^1\| < \|x^* - q_p^2\|) & \Rightarrow x^* \in \omega_1; \\ \text{If } (q_{1-p}^1 < x^* < q_p^2) \ \wedge (\|x^* - q_{1-p}^1\| > \|x^* - q_p^2\|) & \Rightarrow x^* \in \omega_2. \end{array} \tag{1}$$

The reader will observe that the cases are mutually exclusive and that the classification border is: $\frac{q_{1-p}^1 + q_p^2}{2}$.

Case 2: When the quantiloids are overlapping (see Figure 1 on the right):

$$\begin{array}{ll} \text{If } x^* < q_p^2 & \Rightarrow x^* \in \omega_1; \\ \text{If } x^* > q_{1-p}^1 & \Rightarrow x^* \in \omega_2; \\ \text{If } (q_p^2 < x^* < q_{1-p}^1) \ \wedge (\|x^* - q_p^1\| < \|x^* - q_{1-p}^2\|) & \Rightarrow x^* \in \omega_1; \\ \text{If } (q_p^2 < x^* < q_{1-p}^1) \ \wedge (\|x^* - q_p^1\| > \|x^* - q_{1-p}^2\|) & \Rightarrow x^* \in \omega_2. \end{array} \tag{2}$$

In this case, the comparison is based on the distant quantiloids and so the classification border is: $\frac{q_p^1 + q_{1-p}^2}{2}$.

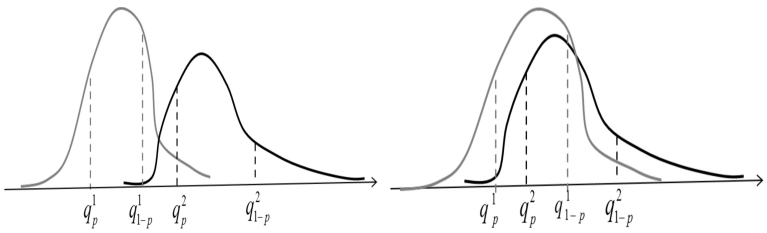

Fig. 1. The AB scheme: (a) When the quantiloids are non-overlapping on the left, and (b) When the quantiloids are overlapping on the right.

The reader will observe that the latter case (Case 2) is the one that uses the so-called "Dual" scenario (please see [8], [9] and [10]), and where the extreme quantiloids are used for the classification as opposed to the quantiloids that are close to the discriminant. In the symmetric cases analyzed in [8], [9] and [10], it is easy to see that the assignments in the so-called "Dual" scenario reduce to those involving comparisons to the quantiloids that are *close to the discriminant*, but where the assignment is to the class *that is the more distant one*. The decision rule for this is given below.

Case 2 (Revised): When the quantiloids are overlapping (again see Figure 1 on the right):

$$\begin{array}{ll} \text{If } x^* < q_p^2 & \Rightarrow x^* \in \omega_1; \\ \text{If } x^* > q_{1-p}^1 & \Rightarrow x^* \in \omega_2; \\ \text{If } (q_p^2 < x^* < q_{1-p}^1) \ \wedge (\|x^* - q_p^2\| < \|x^* - q_{1-p}^1\|) & \Rightarrow x^* \in \omega_2; \\ \text{If } (q_p^2 < x^* < q_{1-p}^1) \ \wedge (\|x^* - q_p^2\| > \|x^* - q_{1-p}^1\|) & \Rightarrow x^* \in \omega_1. \end{array} \tag{3}$$

The difference between the two versions of Case 2 (Eq. (2) and (3)) lies in the assignments made in the last two statements, where they, however, are *done to the non-adjacent classes*. In this case, the comparison is based on the closer quantiloids and so the classification border is: $\frac{q_p^2 + q_{1-p}^1}{2}$. To distinguish between these two scenarios, we shall refer to this version of the "Dual" scenario as the "Swapped Border" scenario.

The cases when the second distribution (for ω_2) is to the left of the first (for ω_1), is shown in Figure 2. Observe that this is identical to the case of the figure on the left of Figure 1, except that the identities of the classes is interchanged.

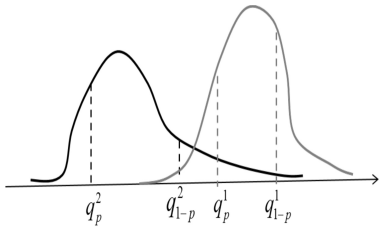

Fig. 2. This figure depicts the case of when the quantiloids do not overlap but when second distribution (for ω_2) is to the left of the first (for ω_1).

There is one additional scenario, and that occurs when there is a huge overlap between the distributions (See Figure 3). The classification decision rule to be used is not that obvious because the classes are highly overlapping. Apart from this, the classification of an unknown sample itself is not just non-obvious, it is actually "meaningless". This case never occurred in our experiments.

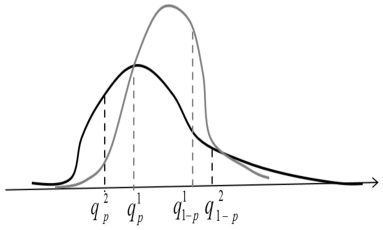

Fig. 3. This figure depicts the scenario when there is a huge overlap between the distributions.

3 The AB Multi-dimensional Clustering

We now consider the extensions of the results in [1] to the multi-dimensional scenario. To explain this, we state that in [1], as explained above, we used the concept of the closest quantile corners in two dimensions. For the multi-dimensional scenario, instead of measuring the distances between the centroids as as done in the Bayesian paradigm, we measure the distances between the quantiloids

3.1 The Quantiloids Used

In the d-dimensional feature space, let $Q^1 = \left[Q_1^1, Q_2^1, ..., Q_d^1\right]$ and $Q^2 = \left[Q_1^2, Q_2^2, ..., Q_d^2\right]$ denote the quantiloids of the distributions (clusters) of $f_1(X)$ and $f_2(X)$ respectively. Q^1 and Q^2 are computed as follows:

- In each dimension, we decide which distribution (cluster) is to the left and which is to the right. To decide this, we exactly follow the principles explained in [1].
- For each of the three cases defined in [1] the elements in the i^{th} dimension of the quantiloid vectors is computed as follows:
  - Case 1: Here $f_1()$ is to the left of $f_2()$. Here we set $Q_i^1 = q_{i,p}^1$ and $Q_i^2 = q_{i,1-p}^2$. In this case, we also have to consider the case when an exception occurs, i.e., when there is a degree of overlap between them. Indeed, if the $f_1()$ and $f_2()$ are close in the i^{th} dimension such that the quantiles overlap, i.e. that $q_{i,p}^1$ is to the right of $q_{i,1-p}^2$, then the point should be classified to ω_1 if it is closer to $q_{i,1-p}^2$ and to ω_2 if it is closer to $q_{i,p}^1$. This corresponds to "Swapped Border" scenario (Case 2 (Revised)) in Section 2.2. With such overlapping quantiles we therefore set $Q_i^1 = q_{i,1-p}^2$ and $Q_i^2 = q_{i,p}^1$.
  - Case 2: Here $f_1()$ is to the right of $f_2()$ and, if the quantiles do not overlap, we set $Q_i^1 = q_{i,1-p}^1$ and $Q_i^2 = q_{i,p}^2$. If the quantiles overlap, we switch the quantiles, as described above, to account for the "Swapped Border" scenario as in Case 1.
  - Case 3: Here we set $Q_i^1 = \frac{q_{i,p}^1 + q_{i,1-p}^1}{2}$ and $Q_i^2 = \frac{q_{i,p}^2 + q_{i,1-p}^2}{2}$. This is the case when the overlap is significant and the classification can be considered to be "meaningless". As mentioned earlier, this case occurs very rarely in the domain of clustering.

The first two scenarios encountered above can be explained in the following figures drawn in two dimensions. In each case, we have plotted the hyper-rectangles defined by the quantiloids. If the overlap is small, the distances are measured from the nearest quantiloids, as seen in Figure 4.

If the overlap is significant, the distances are measured from the fartherest quantiloids (Case 2 of Section 2.2), or equivalently from the "Swapped Border" (Case 2 (Revised)) quantiloids explained in Section 2.2 and shown in Figure 5.

3.2 The Distance Measures Used

Based on the definition of the quantiloids, we are now ready to define two types of distances used in the framework of our AB clustering paradigm. The two types of distance metrics we use are listed below:

- *Data Point to Cluster (DPC) Distance:* Once the quantiloids have been computed following the procedure above, the points Z is classified to ω_1 if Z is closer, in terms of its Euclidean distance to Q^1 than to Q^2. Otherwise, Z is classified to ω_2. The DPC Distance has been used for flat clustering as well as for Top-Down clustering.

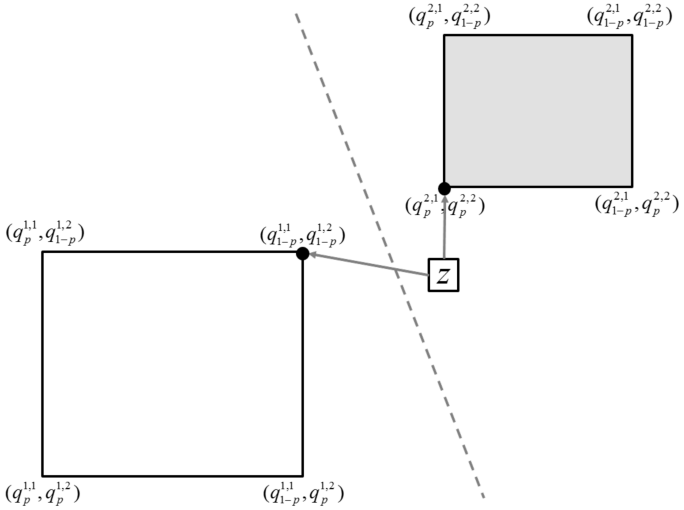

Fig. 4. The case when the multi-dimensional distributions have little overlap. The rectangles representing the quantiloids.

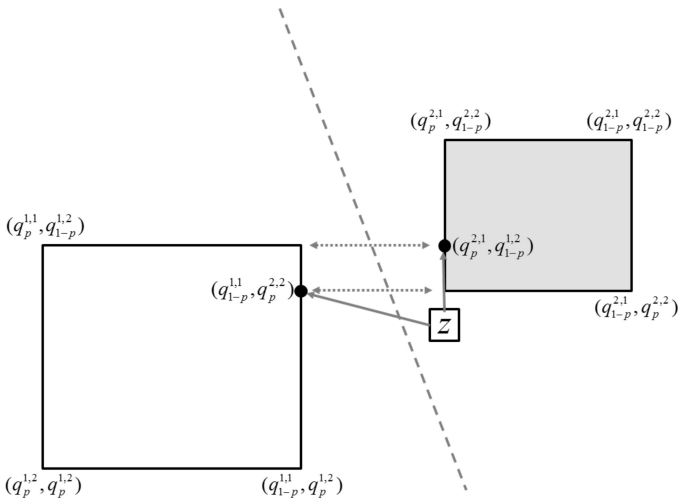

Fig. 5. The case when the multi-dimensional distributions have a large overlap. The rectangles representing the quantiloids. Observe that in this case, we have utilized the "Swapped Border" scenario to compute the quantiloids.

– *Cluster to Cluster (CC) Distance* The same notion can be used to characterize the distance between two clusters. The CC distance between two clusters C_1 and C_2 is the Euclidean distance between their corresponding quantiloids

Q_1 and Q_2. The notion of the CC Distance is usually used for Bottom-Up clustering techniques.

4 Principles of our Hierarchical Clustering

It is well-known that clustering can also be achieved hierarchically, where the scheme is either of a Bottom-Up paradigm or of a Top-Down paradigm. These traditional paradigms can be extended to our AB paradigm by *merely modifying the concept of the distances invoked*, where in the AB scheme, the distance is based on the concept of quantiloids. Thus, in essence, our algorithms follow the classical hierarchical clustering philosophy [11] in all the relevant steps, except that we consider the distances to the qunatiloids rather than the distances to the centroids of the clusters. To explain these, we present the hierarchical AB clustering methods. These are, precisely, the counter-parts of the classical hierarchical clustering methods [11]. The only difference is the way by which we specify the distances, i.e., whether we invoke the DPC or CC distance measures based on the principles of the quantiloids rather than centroids.

4.1 Bottom-Up AB Clustering

A Bottom-Up clustering works with the principle that all the points are individually specified in the d-dimensional space. The points and then gathered together to form clusters, to which the unclassified points are then subsequently added. Thus, the steps of a Bottom-Up AB clustering are described below:

- Compute all pair-wise similarity distances between the different clusters and populate the proximity matrix. The distance between the clusters is merely the Euclidean distance between their corresponding quantiloids.
- Identify the closest clusters in terms of their similarity and merge them into a single cluster. This results in updating the proximity matrix and decreasing its order by unity.
- Repeat the above steps until we obtain the desired (pre-specified) number of clusters.

4.2 Top-Down AB Clustering

A Top-Down clustering works with the principle that all the points are collectively grouped into a single cluster in the d-dimensional space. The most distant points are then separated to be the nuclei of two distinct clusters, and the points closest to these are then included into their respective clusters. Again, in an AB paradigm, the distances are measured in terms of the quantiloids rather than the centroids. Thus, the steps involved in Top-Down AB clustering are described below:

- Start at the top level with all the data points coalesced in a single cluster.
- Use a flat clustering scheme in order to split the cluster.

– Apply the procedure recursively until a termination condition on the depth of the tree is reached or until each data point (singleton) ends up as its own cluster (maximum depth). Usually, one invokes a termination condition which involves the desired (pre-specified) number of clusters.

5 Experimental Results: "Anti–Bayesian" Clustering

In order to test the validity of the concepts proposed in this paper, we conducted numerous experiments on synthetic data. In the interest of space and brevity, we report the salient ones here.

In all our experiments, we used $K = 3$ clusters. All the synthetic data were from multivariate Normal distributions, where we fixed d, the dimension of the space, to be 4.

5.1 Data Generation

We shall first explain how the data points were generated for Normally-distributed distributions. Let $N(\mu_k, \Sigma_k)$ denote a multivariate Normal distribution with an expectation vector μ_k and a covariance matrix Σ_k, where $k = 1, \ldots, K$ (where we are dealing with K clusters). To generate the K distributions, it is crucial that we determine how μ_k and Σ_k $(k = 1, \ldots, K)$ are set.

In our experiments, the expectations, $\{\mu_k\}$ were uniformly spread on the $d-$dimensional cube $[0, D]^d$ where D was chosen such that the clusters were reasonably spread in space to their inter-class overlaps to be minimal. In the experiments for which we report the results, we used $D = 6$.

The covariance matrices Σ_k for each cluster was generated by the following procedure:

- Set the diagonal element to be equal to 1, i.e. the marginal variance was equal to unity in all the dimensions;
- The correlation between each of the variables in the i^{th} and j^{th} dimensions for $i = 1, \ldots, j < i$ was drawn uniformly from the interval $[-\rho_{\max}, \rho_{\max}]$, where $\rho_{\max} < 1$. In the experiments we used $\rho_{\max} = 0.8$;
- We checked if the generated covariance matrix was positive definite. If it was not, we repeated the previous step, above[3].

In what follows, we let n denote the number of samples generated from each underlying cluster in the synthetic data. Thus, the total number of data points generated were $n \cdot K$. In the experiments that we report, we used two values of n, i.e., $n = 20$ and $n = 100$.

[3] With $\rho_{\max} = 0.8$ and for $d = 3$ the matrix were almost always positive definite on the very first attempt. For $d = 5$, on the average, about every third matrix that was generated was positive definite.

5.2 Quantile and "Distance" Estimations

Since we constantly need to estimate the quantiloids, we opted to achieve this using the corresponding quantiles in each of the projected dimensions. This was done non-parametrically and parametrically as below:

- Non-parametrically (referred to in the columns titled "AB Non-parametric" in the tables below): This was achieved in a manner similar to the work presented in [1].
- Parametrically: This was achieved by assuming normality Here we estimated the corresponding μ and σ and computed the respective quantiles from the Normal distributions (referred to in the columns titled "AB Parametric" in the tables).

The corresponding "distance" estimations for the experiments done were achieved as below:

- Row titled "Bottom up": These represent the classical version where all the inter-point distances are computed. The distances were computed between the centroids in case of the Bayesian clustering, and between the corresponding quantiloids in the case of the AB clustering.
- Row titled "Bottom-Up Distance UD (uni-dimensional)": In this case, we sorted the data by the first dimension and repeatedly merged the closest points in this dimension. This approach was a simplification of the general approach where we should have considered all the dimensions. The present approach required less computations. We expected that such a simplification would result in a reduced accuracy as we only relied on the first dimension of the data for executing the clustering, and this was, indeed, our experience.
- Row titled "Top-Down": In this case, the points were repeatedly split in two using k-means and the AB analog ([1]).
- Row titled "Top-Down Distance UD (uni-dimensional)": In this case, the data was sorted by the first dimension and repeatedly split in such a way that the distance (Bayes or AB) between the clusters was as large as possible. Again, this approach required less computations than the "Top-Down" one. As before, it is reasonable to expect such a simplification to result in a reduced accuracy. This was, indeed, the case.

5.3 Evaluation of Clustering Performance

The better a clustering algorithm performs, the better we expect that the following is satisfied: If two points are in the same clusters for the true clusters described by the "state of nature', they should ideally also be in the same cluster in the results obtained from the clustering algorithm. Conversely, if two points are *not* in the same clusters for the true clusters, they should not be in the same cluster in the results of the clustering algorithm either. As a measure of clustering performance, we measured the portion of pair of points satisfying this agreement between the true clusters and the clusters from the algorithm.

5.4 Clustering performance

Tables 1 and 2 show results for the different clustering algorithms when $n = 20$ and $n = 100$, respectively. Comparing the non-parametric and parametric AB

	Bayes	AB Non-parametric	AB Parametric
Flat clustering	0.098 (0.094, 0.103)	0.098 (0.094, 0.102)	0.099 (0.094, 0.103)
Bottom-Up	0.21 (0.201, 0.219)	0.277 (0.266, 0.288)	0.275 (0.265, 0.286)
Bottom-up Distance UD	0.423 (0.415, 0.43)	0.434 (0.427, 0.442)	0.443 (0.435, 0.45)
Top down	0.13 (0.125, 0.135)	0.132 (0.127, 0.137)	0.13 (0.125, 0.135)
Top-Down Distance UD	0.335 (0.329, 0.341)	0.369 (0.363, 0.375)	0.358 (0.352, 0.365)

Table 1. The clustering errors of the various methods with $n = 20$.

	Bayes	AB Non-parametric	AB Parametric
Flat clustering	0.079 (0.064, 0.094)	0.089 (0.072, 0.107)	0.082 (0.065, 0.099)
Bottom-Up	0.247 (0.203, 0.291)	0.322 (0.268, 0.375)	0.369 (0.311, 0.428)
Bottom-up Distance UD	0.479 (0.451, 0.507)	0.481 (0.451, 0.512)	0.504 (0.476, 0.532)
Top-Down	0.108 (0.087, 0.129)	0.097 (0.08, 0.114)	0.126 (0.104, 0.148)
Top-Down Distance UD	0.346 (0.32, 0.372)	0.364 (0.342, 0.385)	0.334 (0.311, 0.358)

Table 2. The clustering errors of the various methods $n = 100$.

approaches, we do not observe any (significant) differences in the results showing that both approaches perform about equally well. Comparing the AB approaches to the Bayesian approach we observe that the two methods perform about equally well for all the methods except for the Bottom-Up approach in which the Bayes perform a little better. We also observe that the uni-dimensional approaches (rows three and five) perform far poorer than the other methods documenting that such a uni-dimensional approaches are not satisfactory. Overall we see that the Anti-Bayesian paradigm performs very well in clustering data and competitive to the Bayesian paradigm.

6 Conclusion

In this paper, we have considered an "Anti-Bayesian" (AB) paradigm for clustering. All of the reported clustering algorithms (except the one reported in [1]) operate on Bayesian principles, (where the Bayesian principle corresponds to assigning the unlabelled samples to the cluster whose mean (or centroid) is the closest). Our aim here has been to see if the "Anti-Bayesian" (AB) classification philosophy, introduced recently by Oommen and his co-authors, can be extended into the domain of clustering. The AB principle involves classification

based on the non-central *quantiles* of the distributions, which involves utilizing the information resident in the outlier samples.

In this paper, we have extended the first-reported AB clustering methods proposed in [1]. This paper has extended the results of [1] in many directions. Firstly, we have generalized our previous AB clustering [1], initially proposed for handling uni-dimensional and two-dimensional spaces, to arbitrary d-dimensional spaces using their so-called *"quantiloids"*. Secondly, we have extended the AB paradigm to consider how the clustering can be achieved in hierarchical ways, where we have analyzed both the Top-Down and the Bottom-Up clustering options. The AB paradigm can also use an anti-Naïve-Bayesian *computational* mechanisms. The paper contains the results of extensive experimentation that demonstrate that our clustering achieves results competitive to the state-of-the-art flat, Top-Down and Bottom-Up clustering approaches.

In the future, we envisage an ambitious goal of devising an AB clustering method based on applying majority voting on the decision made in each dimension of the quantile vector.

References

1. Hammer, H.L., Yazidi, A., Oommen, B.J.: A novel clustering algorithm based on a non-parametric "anti-bayesian" paradigm. In: Current Approaches in Applied Artificial Intelligence - 28th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2015, Seoul, South Korea, June 10-12, 2015, Proceedings. (2015) 536–545
2. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Dats. Prentice Hall, Englewood Cliffs, New Jersey, USA (1988)
3. Xu, R., Wunsch, D.: Survey of clustering algorithms. Trans. Neur. Netw. **16**(3) (2005) 645–678
4. Ankerst, M., Breunig, M.M., Kriegel, H.P., Sander, J.: Optics: Ordering points to identify the clustering structure, ACM Press (1999) 49–60
5. Ester, M., Kriegel, H.P., S, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise, AAAI Press (1996) 226–231
6. Murtagh, F., Contreras, P.: Methods of hierarchical clustering. CoRR **abs/1105.0121** (2011)
7. Sibson, R.: SLINK: An optimally efficient algorithm for the single-link cluster method. The Computer Journal **16**(1) (1973) 30–34
8. Thomas, A., Oommen, B.J.: The fundamental theory of optimal "anti-bayesian" parametric pattern classification using order statistics criteria. Pattern Recognition **46**(1) (2013) 376–388
9. Thomas, A., Oommen, B.J.: Order statistics-based parametric classification for multi-dimensional distributions. Pattern Recognition **46**(12) (2013) 3472–3482
10. Oommen, B.J., Thomas, A.: Anti–Bayesian parametric pattern classification using order statistics criteria for some members of the exponential family. Pattern Recognition **47**(1) (2014) 40–55
11. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM computing surveys (CSUR) **31**(3) (1999) 264–323