TRIED Etude de Cas 2: Signature Biometrics

Carl Robinson & Nelson Fernandez **21st February 2018**

Dataset

- Dataset name: MCYT-100 database
- Creator: Autonomous University of Madrid
- Data: 100 candidates, 25 samples of genuine signatures per candidate (2500 total signature files)
- Only real signatures were used - the dataset also contains forged signatures but these were discarded.

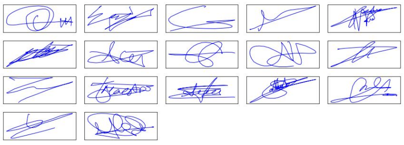
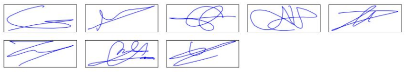
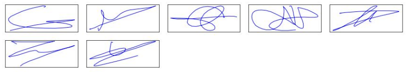
Measuring the Influence of the Number of Gaussians

- Each person has 25 signatures with unique characteristics. The signatures of an individual are very similar, but do vary to some extent, so we model a person's signatures using a mixture of Gaussian distributions.
- Each Gaussian component represents a probability distribution of a person's signature. We create a Gaussian Mixed Model using by combining these Gaussians.
- First we compute the complexity of each signature of each person by calculating the entropy of all its points. This is done using the determinant of the covariance matrix. The mean of the 25 signature complexities is then calculated for each person.
- We use k-means to cluster the people into 3 clusters, using these mean complexity values.
- We can visualize the signatures of the people in the different clusters, in order to see how complex the signatures are in the 3 clusters. Our expectation is that each cluster will contain signatures of a comparable complexity. We choose k=3 to cluster signatures with low, medium and high complexities.
- We then repeat the above process using a different number of Gaussians to characterise a person's signatures. By visualising the signatures in the resulting clusters we can observe how changing the number of Gaussians affects the complexity of signatures that get clustered together. Our expectation/hope is that a greater number of Gaussians, the finer the level of differentiation, and therefore the more effective the clustering will be at grouping signatures with a similar complexity together.

Visualisation of K-means Clusters

In the following table, we show one signature from each person in each cluster (calculated using the mean complexities of their 25 signatures). This allows us to visually assess the quality of the clustering in terms of the complexity of the signatures.

Note: all ‘pen ups’ (where tablet pressure = 0) were removed from these images for ease of interpretation, but the X/Y coordinate values of pen ups are still used in the calculation of complexity.

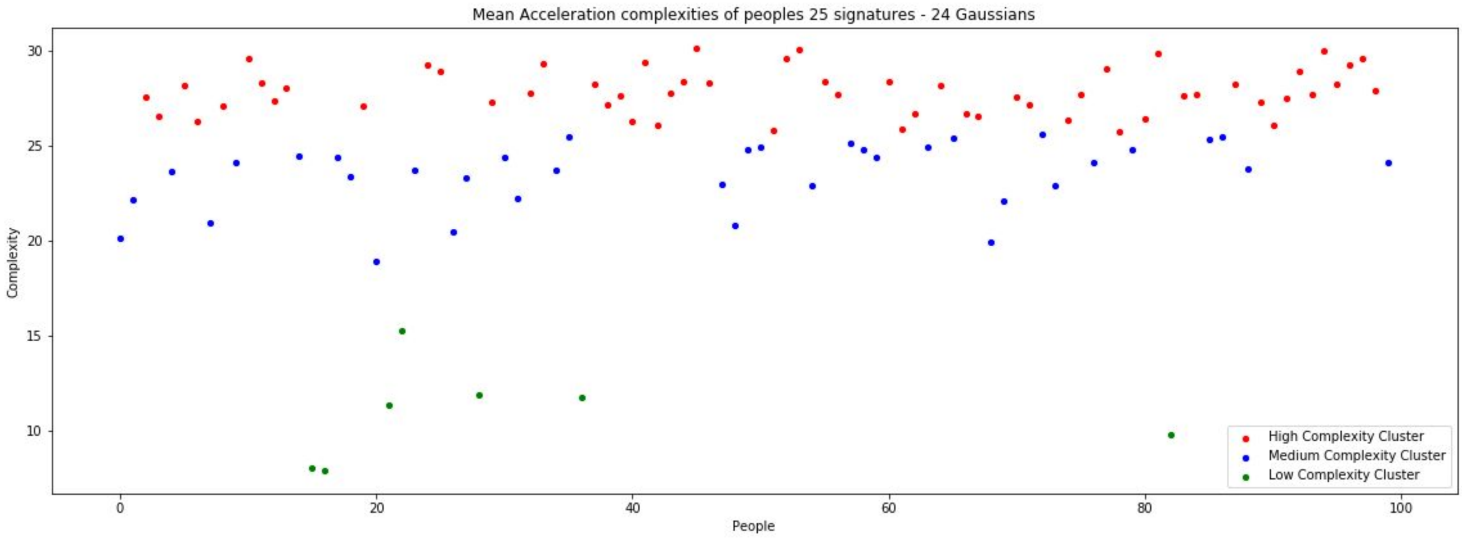
	4 Gaussians	**8 Gaussians**	**24 Gaussians**
Cluster 1	 Complexity: [23.653, 28.260]	 Complexity: [16.347, 24.724]	 Complexity: [2.060, 17.041]
Cluster 2	 Complexity: [28.268, 30.548]	 Complexity: [24.836, 28.982]	 Complexity: [17.253, 25.612]
Cluster 3	 Complexity: [30.550, 34.526]	 Complexity: [28.99, 34.10]	 Complexity: [25.615, 32.071]
Silhouette Score	**0.553**	**0.564**	**0.603**

- We see that the number of Gaussians has a significant effect on the quality of the clusters.
- With 4 Gaussians, cluster 1 contains some medium complexity signatures, and cluster 3 contains some low complexity signatures.
- With 8 Gaussians, cluster 1 contains only very simple signatures, whereas clusters 2 and 3 seem to contain a similar mix of medium to high complexity signatures.
- With 24 Gaussians we achieve the best clustering, with cluster 2 clearly containing less complex signatures than those in cluster 3, but more complex than those in cluster 1.

- Our hypothesis that a greater number of Gaussians would lead to a more effective clustering seems to be validated. With 24 Gaussians, signatures with a similar complexity are grouped together, with very few exceptions.

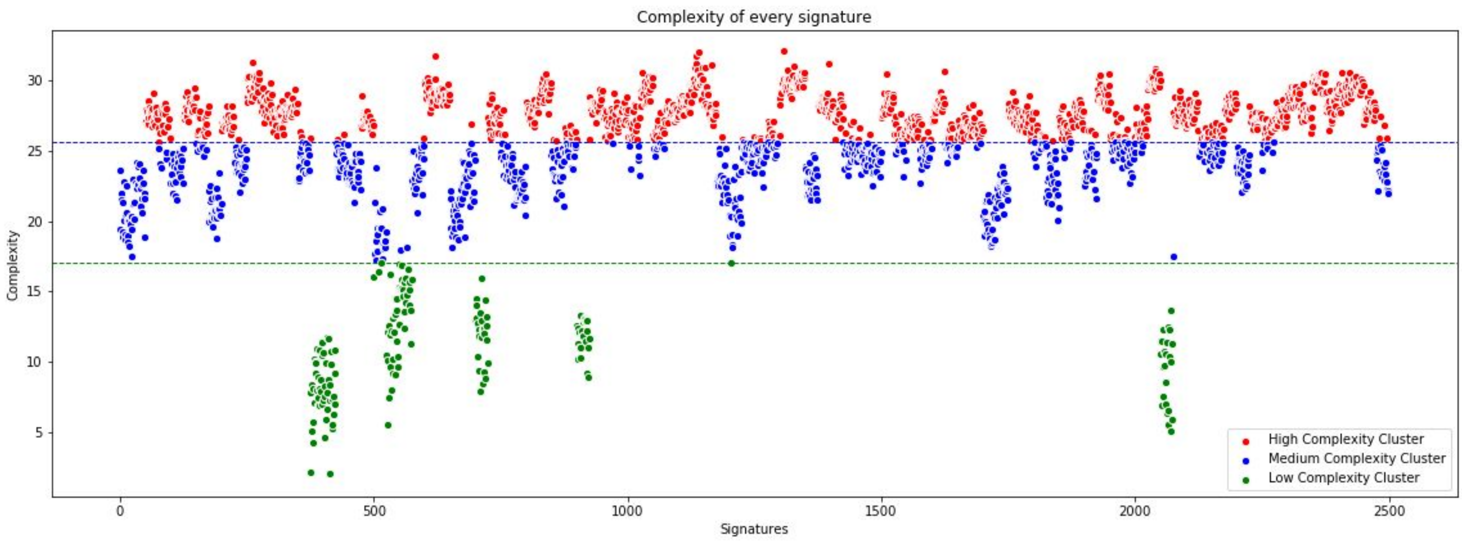
Analysis of Complexity Clustering

Given the complexity measures for the 25 genuine signatures with NG=24 for all users, a K-means (K=3) clustering was performed on the mean complexity values for each user:



We can see that the mean complexities of each user are much closer to each other in the high complexity cluster than they are in the other two clusters.

A K-means (K=3) clustering was then performed on the complexity values at at the level of the signatures (on 2500 complexity values):



The high complexity cluster has 1440 signatures in, the medium complexity cluster has 884, and the low complexity cluster has only 216. Clearly visible in the second graph is the vertical arrangement of signatures from the users in the 3 clusters. Each vertical grouping is a single user/small group of users, as indicated by the signature indexes on the x axis (a single user's signatures would have consecutive signature indexes, therefore appear together in horizontal order on the graph). The vertical spread indicates the high variability in complexity of the signatures produced by a user i.e. a single user's signatures vary in complexity due to inconsistencies in the writing style.

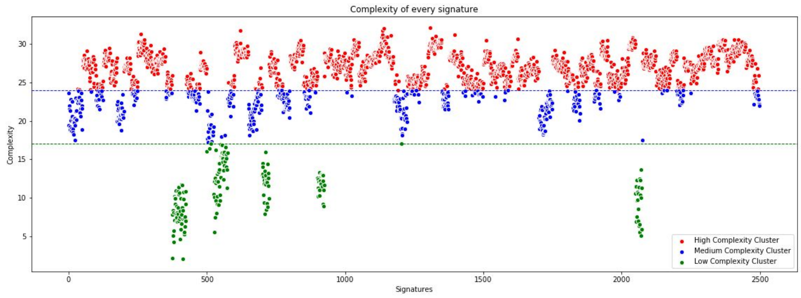
We also see that signatures in the low complexity cluster (green) have the greatest variability in complexity. Less variation is seen in the medium complexity cluster (blue), and the least variability of all in the high complexity cluster (red). This is important, as a biometrics application may find it more difficult to identify or verify low complexity signatures, as the high inter-class variance of this group would reduce the confidence in discerning one person's signature from another person's.

Silhouette Score

The Silhouette Score is a method of interpretation and validation of consistency within clusters of data. It is calculated using the mean intra-cluster distance (a) and the mean nearest-cluster distance (b) for each sample. The Silhouette Coefficient for a sample is (b - a) / max(a, b), and is a value that ranges between 1 (best) and -1 (worst). Values near 0 indicate overlapping clusters. K-means clustering (k=3), achieved a good score of **0.603**.

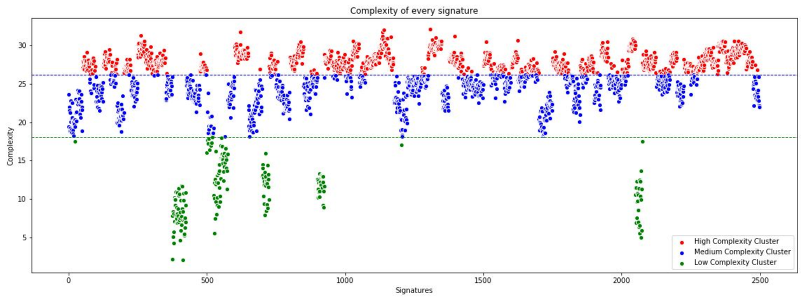
Comparison of Clustering Methods

Agglomerative Hierarchical Clustering (AHC)



- The decision boundary between the low and medium complexity clusters is identical to that found with K-means. Therefore the size of cluster 1 remained unchanged.
- The decision boundary between the medium and high complexity clusters is different, with AHC choosing a lower complexity level as the boundary (23.975 versus 25.612). This increased the size of the High complexity cluster to 1847 signatures, while it decreased the medium cluster to 477.
- The cophenetic correlation metric is a measure of how faithfully the tree represents the dissimilarities among observations. The clustering of the full dendrogram achieved a very good score of 0.870.
- AHC clustering (k=3), achieved a silhouette score of **0.589.**

Bayesian Gaussian Mixed Model (BGMM)



- A Bayesian Gaussian Mixture Model with a Dirichlet process prior fit allows us to adjust the level of detail at which to model.
- BGMM produced a roughly similar clustering to K-means. The decision boundary between the low and medium complexity clusters is slightly higher to that found with K-means. Therefore the size of cluster 1 increased slightly.
- The decision boundary between the medium and high complexity clusters is slightly higher too.
- BGMM creates a more even distribution of signatures between the two clusters: 1286 and 1030 signatures respectively, with the low complexity cluster having 184.
- BGMM clustering (k=3), achieved a silhouette score of **0.593.**

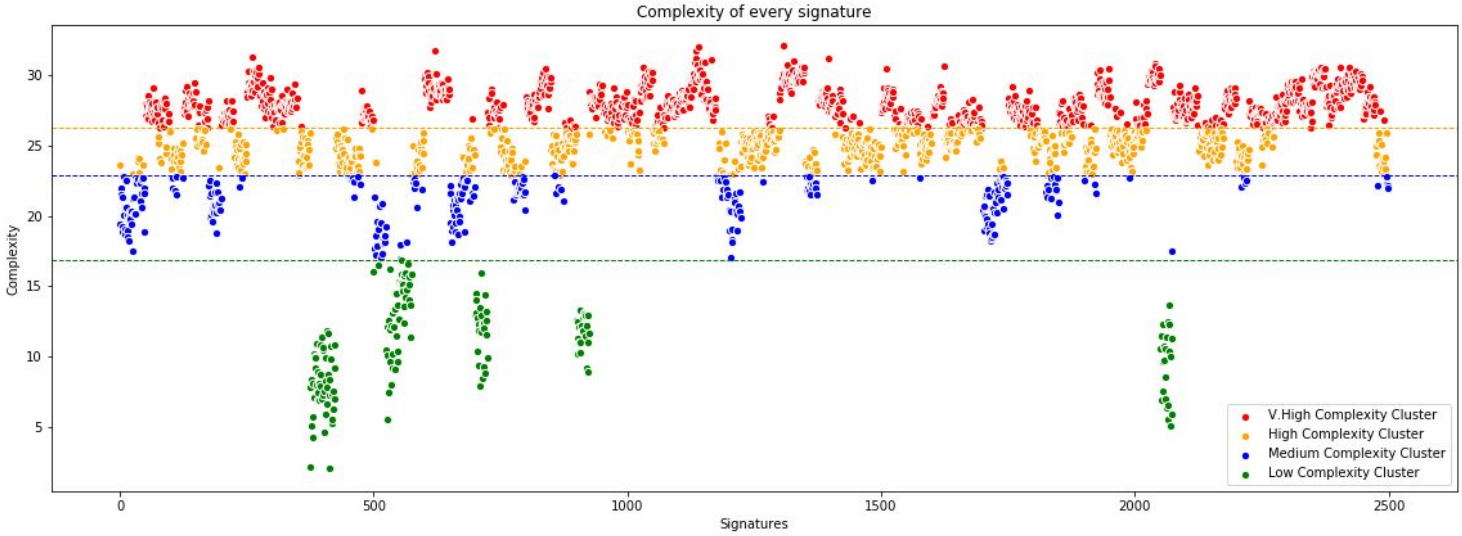
Conclusions

- The three clustering methods produced somewhat similar clusterings, which could be expected given the very simple one-dimensional data. K-means had the highest silhouette score, the others were very similar.
- It can be noticed that all clustering methods produced a large sized cluster for high complexity signatures, a small sized cluster for low complexity signatures. The size of the medium complexity cluster varied.
- It may be possible to refine the large clusters by splitting further, therefore we will explore varying the number of clusters next. This will help to further subdivide the signatures into even more homogenous groups.

Varying the Number of Clusters

We will explore the effect of increasing the number of clusters when using BGMM. Complexities obtained with 24 Gaussians will again be used.

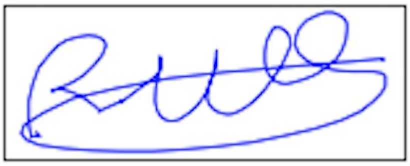
BGMM K=4



With 24 gaussians we can see more clearly that some people's signatures are split across a cluster. The algorithm has kept the large majority of a person's signatures in a single cluster, but due to the choice of k=4 which has created an additional decision boundary, some signatures of a person have been assigned to an adjacent cluster.
BGMM clustering (k=4), achieved a silhouette score of **0.569**, slightly lower than when k=3.

Low complexity (173): 	**Medium complexity (295):** 
High complexity (747): 	**V.High complexity (1285):** 

- There is a clear pattern of increasing complexity across the four clusters, and each cluster is fairly distinct.
- There are still a small number of signatures that appear to be classified, being either significantly more simple or more complex than the other members of their cluster. This may be due to the overall shape of the signature being similar to the other more complex ones, or it could be due to having a large number of pen up points that we're not seeing in the images. For instance, the signature to the right is clearly not of 'very high' complexity, as its cluster would suggest.



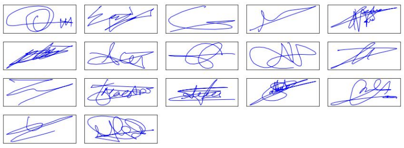
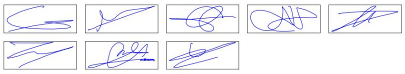
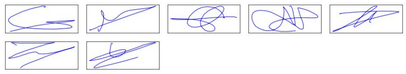
BGMM K=5

- BGMM clustering (k=5), achieved a silhouette score of **0.552**, lower than when k=3 and k=4.
- As it is difficult to visually discern differences between the signatures in 5 or more clusters, and as the silhouette score is decreases as k increases, we chose **k=3** as the optimal number of clusters.

Complexity of Acceleration Values

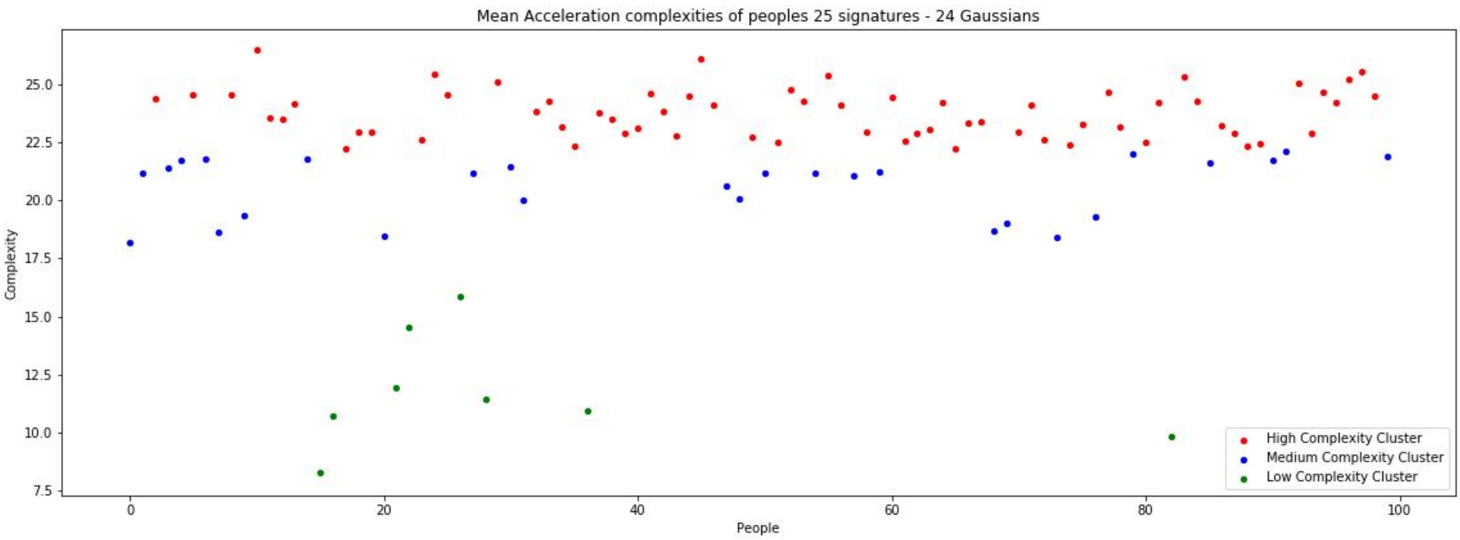
It is possible to calculate the acceleration between each point using the raw X and Y coordinate data provided by the writing tablet. Here we will explore whether this acceleration data can be used to effectively cluster signatures by complexity. We must recalculate the complexities based on acceleration, has it is a completely new signal.

In the following table, we show one signature from each person in each cluster (calculated using the mean complexities of their 25 signatures). However, since we have discovered that one person's 25 signatures do not necessarily belong to the same cluster, this means that these particular signatures shown below may not actually belong to the cluster stated.

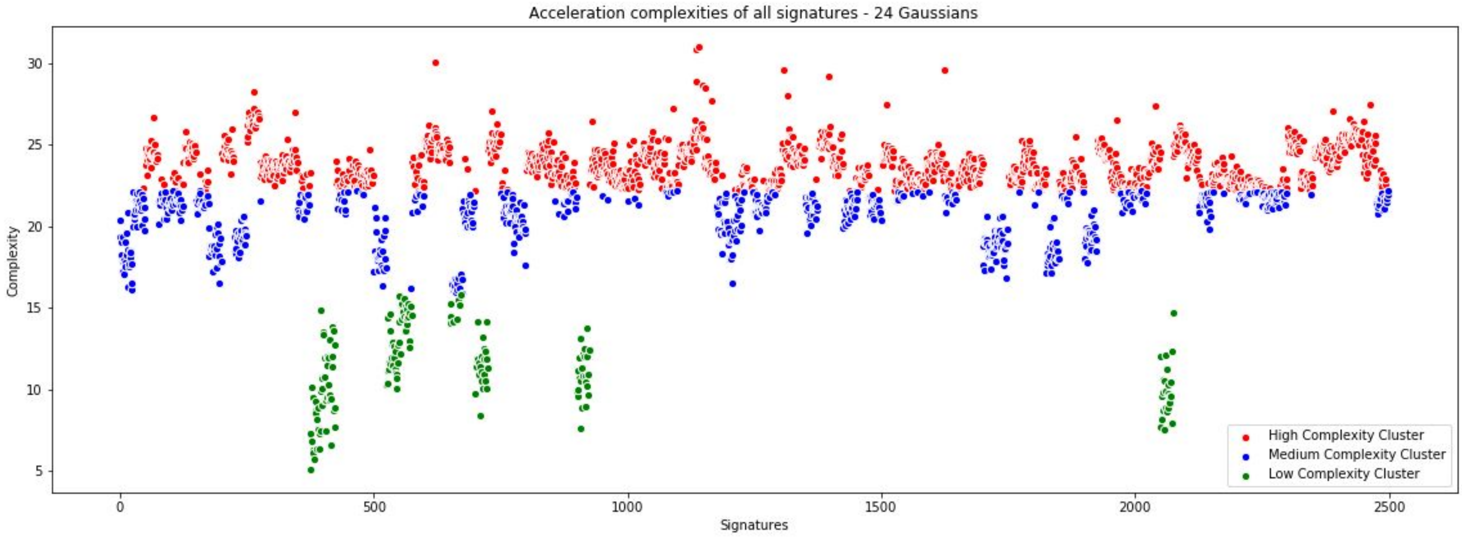
	4 Gaussians	8 Gaussians	24 Gaussians
Cluster 1	 Complexity: [18.485, 23.367]	 Complexity: [14.575, 21.646]	 Complexity: [5.1245, 15.843]
Cluster 2	 Complexity: [23.377, 25.803]	 Complexity: [21.665, 24.928]	 Complexity: [15.924, 22.176]

Cluster 3	 Complexity: [25.805, 32.407]	 Complexity: [24.933, 31.788]	 Complexity: [22.178, 30.978]
Silhouette Score	**0.551**	**0.554**	**0.564**

- With 4 and 8 Gaussians, the clusters are of poor quality, with the medium and high complexity clusters being fairly indistinguishable. With 4 Gaussians, even the low complexity cluster is polluted with some reasonably complex signatures. This shows that 4 or 8 Gaussians is insufficient to fully represent all the complexity of a signature.
- With 24 Gaussians, the pattern of increasing visual complexity is still clearly seen; the low, medium clusters and high are now distinct in their visual complexities. However, a few signatures with relatively simple shapes have still been classified as high acceleration complexity.
- Increasing the number of Gaussians further (e.g. 36) may help to correctly classify these, but having too many Gaussians also risks suffering overfitting the signatures, which would also result in a poor clustering. A better solution may be to use acceleration in conjunction with another variable such as the coordinates.



Using acceleration instead of coordinates has still produced a clear separation of people, with complexities ranging from ~8.0 up to ~27.0. This shows that much of the variance between signatures is expressed by the way a person writes their signature (the acceleration) as it is by the shape of the signature itself (x y coordinates).

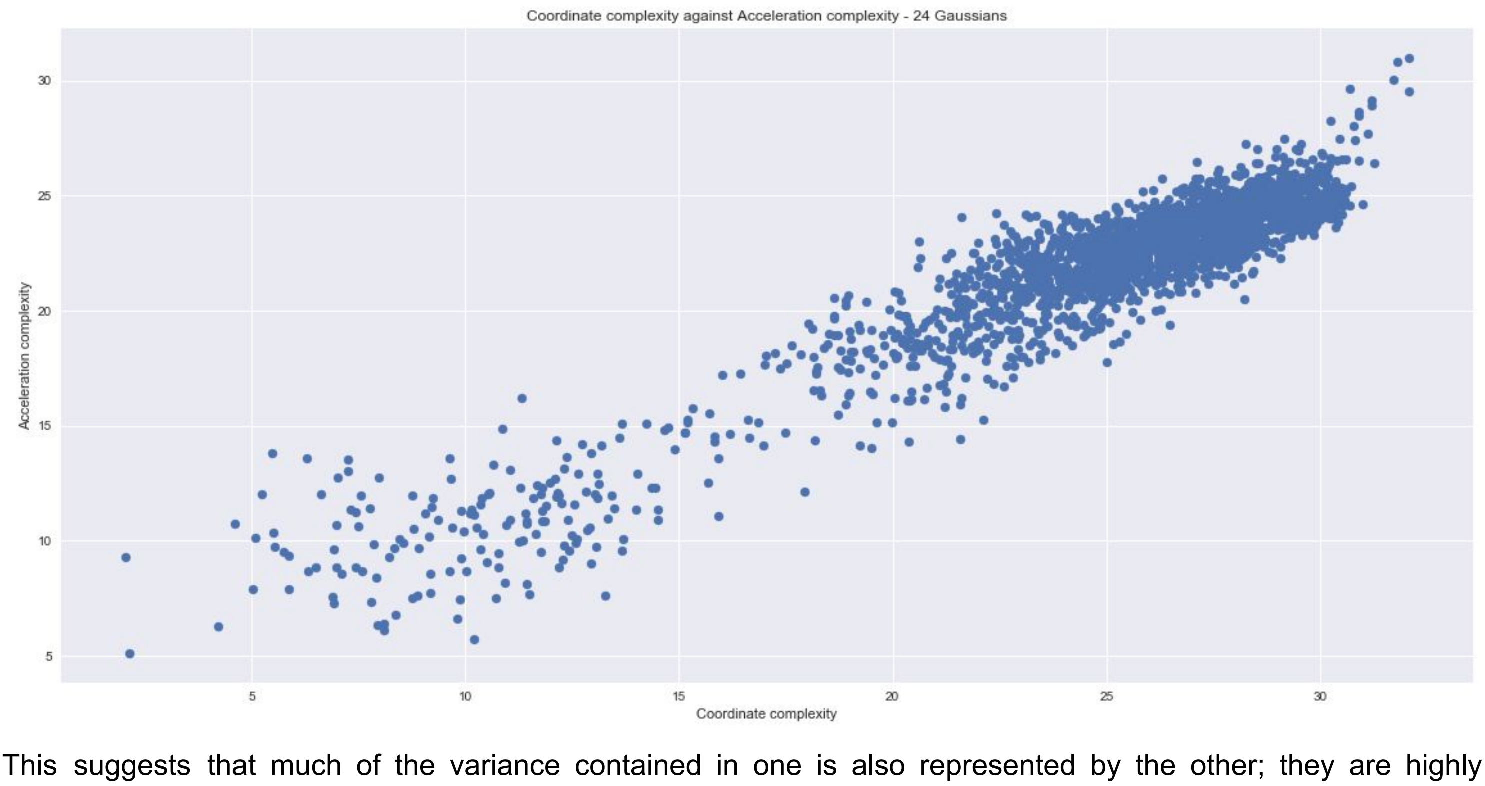


Again, we can see that some people's signatures are split across a cluster, but most people's signatures all or mostly lie within one single cluster. We can see that there are a number of outliers that have especially high complexity, so we will examine these more closely below. Most importantly, we see that there is much greater intraclass variance in the low complexity signatures than in the higher complexity signatures. Here, K-means clustering (k=3) achieved a silhouette score of **0.564**, which is only slightly less than with coordinate complexities.

Overall, this result shows that acceleration is capable of separating signatures into clusters well, but it cannot be considered a substitute for visual complexity on its own. This is important from an application development standpoint, as users of electronic signature technology may be advised to choose 'complex' signatures, by which most people would understand to mean 'visually complex'. Complexity of acceleration would be a foreign concept to most people, and therefore less useful as way for most users to ensure they chose secure signatures that were difficult to forge.

Relationship between Coordinate complexity and Acceleration complexity

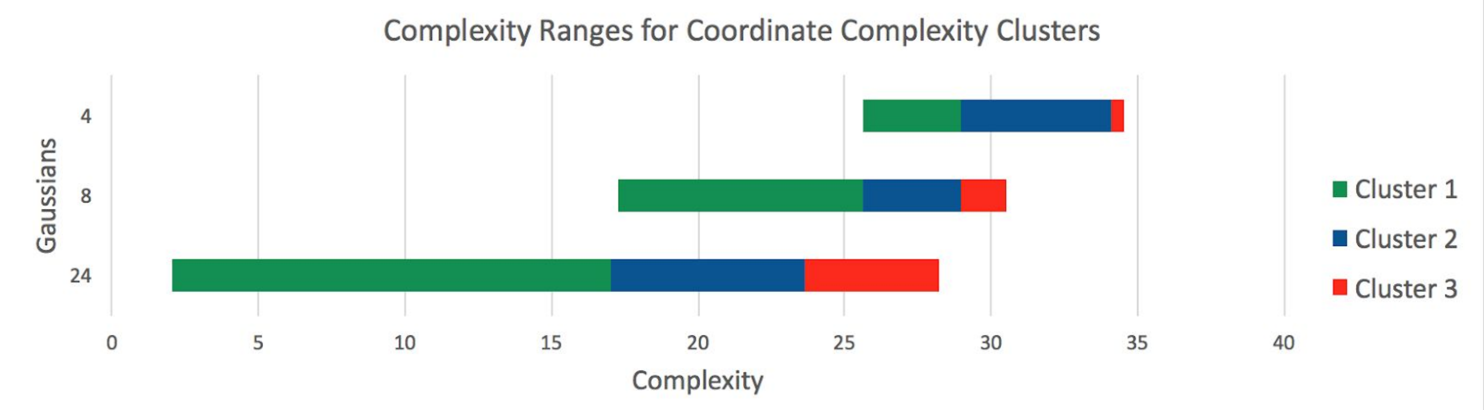
Plotting the coordinate and acceleration complexities shows there is a clear linear relationship between the two:



This suggests that much of the variance contained in one is also represented by the other; they are highly correlated. Using both variables to cluster the signatures would therefore be unlikely to yield significantly improved separation.

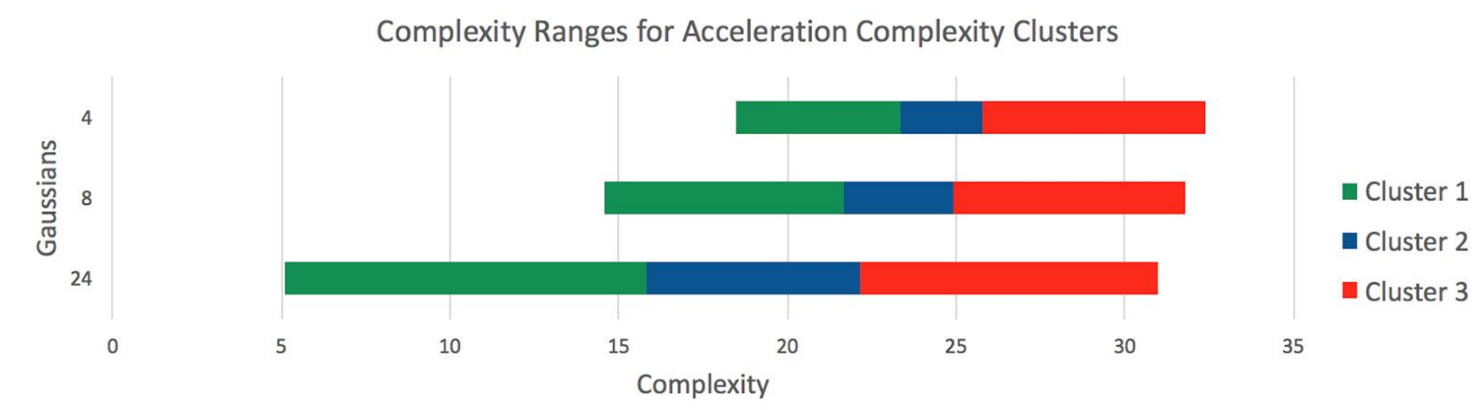
Comparison of Complexity Ranges

As well as a visual analysis of the signatures in each cluster, it is helpful to analyse the magnitude and range of the quantitative complexity values in each cluster.



	4 Gaussians	8 Gaussians	24 Gaussians
Cluster 1	[23.653, 28.260]	[16.347, 24.724]	[2.060, 17.041]
Cluster 2	[28.268, 30.548]	[24.836, 28.982]	[17.253, 25.612]
Cluster 3	[30.550, 34.526]	[28.99, 34.10]	[25.615, 32.071]

We see that the complexity values produced using 4 Gaussians are both higher in value and occur in a much tighter range (23.7 to 34.5) than those produced with 24 Gaussians (2.1 to 32.1). This helps to explain why the clustering algorithms find it more difficult to correctly cluster the signatures, as fewer Gaussians (that explain the structure of the data) overlap more, resulting in higher complexities overall, and are therefore less effective at separating the signatures.



	4 Gaussians	8 Gaussians	24 Gaussians
Cluster 1	[18.485, 23.367]	[14.575, 21.646]	[5.1245, 15.843]
Cluster 2	[23.377, 25.803]	[21.665, 24.928]	[15.924, 22.176]
Cluster 3	[25.805, 32.407]	[24.933, 31.788]	[22.178, 30.978]

As with the coordinate complexities, we see that the complexity values produced using 4 Gaussians are both higher in value and occur in a tighter range than those produced with 24 Gaussians. While 24 Gaussians is still the best, the number of Gaussians has less effect on the magnitude and ranges of the acceleration complexities. This explains why the acceleration complexity clusters are less homogenous than the coordinate clusters.

Outliers and the Effect of Pen-ups

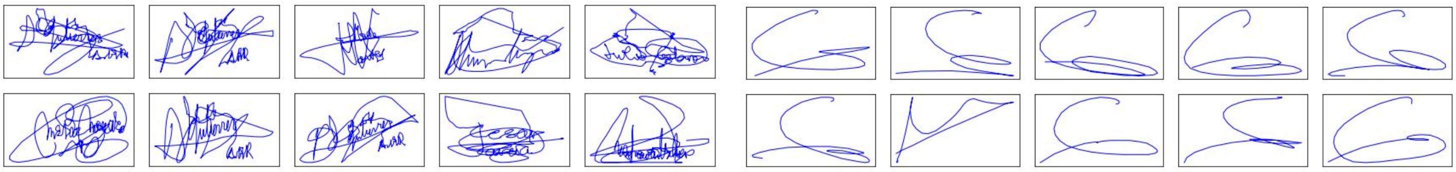
Outliers

The following signatures are those with the highest and lowest acceleration complexity, with pen-ups hidden:

- We see straight away that the outliers are extremely simple or complex in their shape, even when using acceleration complexity to cluster.
- We can also see that the signatures with the highest acceleration complexity are comprised of many sharp angles. These shapes reveal that many changes of direction and speed are the reason for their high acceleration complexity.
- In contrast, the most simple signatures in terms of acceleration complexity are comprised of a few very smooth flowing lines. These lines will have long sequences of gentle changes in speed and acceleration as a result of their shape, and as such, low acceleration complexity.

The Effect of Pen-ups

Here are the same 10 most complex and least complex signatures, but with the pen-ups visible this time:



- By comparing these with the previous figure, we immediately see that the high complexity signatures have many pen-up lines added, whereas the low complexity signatures have very few or no additional lines.
- The pen-up lines that have been added are also of a very 'jerky' nature, giving them high acceleration complexity themselves.
- Pen-ups can be considered a quality issue with the sensor (the writing tablet), as these are unintended consequences of the sensor that would be invisible to the end user of a verification system. A user may be advised to use a complex signature, unaware that this signature would include many more pen-ups, and therefore have a complexity even higher than intended. This may produce undesirable or misunderstood consequences during operation.
- In summary, we have found that pen-ups further contribute to both the spatial and acceleration complexities of signatures, and help to explain the difference in complexity between visually simple and complex signatures.

Relationship between Complexity and Dissimilarity

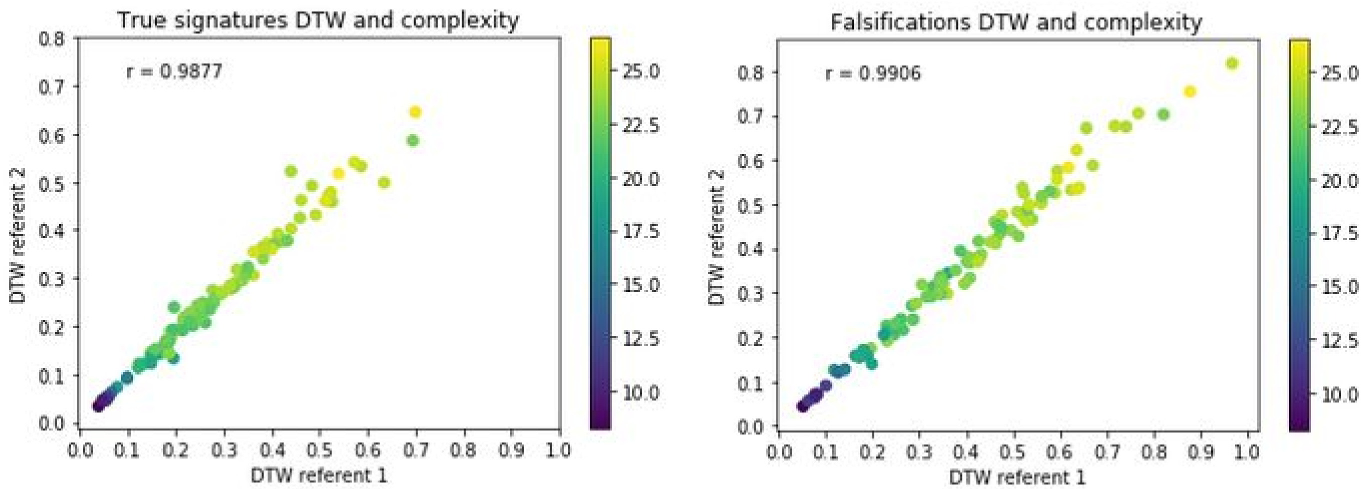
A key factor on signature verification is how susceptible signatures are to counterfeit. As a matter of fact, skilled forgers can imitate complex patterns to violate even the most sophisticated verification systems. However, the quality of the counterfeit is related not only to forger's skills but to the overall complexity of the signature. Thus, less complex signatures are more susceptible to be successfully imitated compared to complex ones.

In this part of the study, we propose a methodology to evaluate the complexity of a signature along with a measure of dissimilarity: Dynamic Time Warping (DTW).

DTW is a method that allows us to compute an elastic distance between two temporal series that do not necessarily have the same length. The advantage of this technique is its capacity to deal with local variations, offsets and outliers, making it an ideal choice for compare highly variable handwritten signatures.

In general, high DTW values suggests a high degree of dissimilarity, meaning the two signatures do not match, whereas low DTW values mean the opposite. As such, DTW is widely used as an acceptance-rejection index in banking and other signature verification applications.

The proposed methodology is the following: For each person we take two of their genuine signatures as valid references, then we compute DTW between these two genuine references and their other twenty-three genuine signatures, and also with the twenty-five forged signatures. Mean DTW values are calculated for genuine and forged signatures, which are then taken as the indices of dissimilarity. DTW values are normalized in a common scale from 0 to 1 for comparison purposes, and presented as follows:



The left figure shows the normalized mean DTW when comparing the genuine signatures against reference #1 (x axis) and reference #2 (y axis), the color bar indicates the complexity of the signature calculated using a 24 Gaussian GMM model. The figure shows a linear relationship between DTW genuine vs reference #1 and genuine vs reference #2. This indicates that DTW values are almost equal when comparing one genuine signature against both references. This behaviour is expected by assuming high quality references and the signer is able to reproduce his/her signature in a reasonably consistent manner.

The complexity values tell another story. Firstly, less complex signatures present low DTW values, suggesting that the signer is able to repeat his/her signature without any major difficulties. Our previous findings showed that low complexity signatures have a high degree of variability in complexity. Therefore we can say that despite this variability, the DTW values remain low for low complexity signatures.

On the other hand, highly complex signatures have higher DTW values, indicating that the intricate patterns of the signature are difficult to replicate exactly, even by the signature's owner. We also see that the high complexity DTW values (shaded yellow) are more spread out, indicating higher variance in DTW values. This is likely due to the increased intraclass variability of a person's high complexity signature. In our previous results, we found that highly complex signatures had less intraclass variation in complexity than their low complexity counterparts. Here we find that in addition to this low variability in complexity, high complexity signatures also have the additional advantage of being difficult to reproduce exactly and consistently.

The right figure shows the results obtained by applying the same methodology to falsified signatures. The figure displays a similar pattern to the previous one. However, the DTW points are spread out in a wider arc, indicating that the difference in DTW values between a forged signature and the genuine reference signatures is more variable. This is to be expected given that a signature forger will be less adept at consistently reproducing another person's signature.

It is important to note that low complexity signatures, both original and counterfeits, present almost the same DTW values in low complexity ranges (from 0 - 12.5), and this difference becomes more pronounced as the complexity increases. This result suggests that low complexity signatures are easy to imitate, as the forgers are able to reach a similar DTW value.

In conclusion, a DTW-based validation system could be consistently bypassed by forgers when the complexity of the genuine signature is low.