Supplementary Materials

We present the supplementary materials to show the efficiency and effectiveness of our enhanced usage of cPCA (ccPCA). In Sect. A, we perform a performance comparison of the original cPCA implementation and ours. This section showcases the efficiency of our implementation and parallel calculation. Sect. B shows the comprehensive experimental results of ccPCA with different values of the contrast parameter α to demonstrate the effectiveness of our automatic selection of α. Lastly, in Sect. C, we provide a comparison of the features' contributions obtained with LDA, PCA, and ccPCA.

A Performance of cPCA with the Automatic Selection of the Best Contrast Parameter α

We present the performance comparison of the original cPCA implementation with the semi-automatic selection of α[1] (Sect. 4.1), our cPCA implementation with the automatic selection of α, and the multi-thread version of our algorithm[2], as described in Algo. 1 (Sect. 4.2).

A.1 Experimental Settings

As an experimental platform, iMac (Retina 5K, 27-inch, Late 2014) was used. It has 4 GHz Intel Core i7 (four cores and eight threads) and 16 GB 1,600 MHz DDR3. Each result below shows the average completion time in five executions. To run the experiment, we generate datasets that contain different numbers of data points n (from 100 to 100,000 data points) with different numbers of dimensions d (from 10 to 1,000). All values are randomly assigned in the $[-1:1]$ range. Also, we set $q = 40$ (q: the number of candidate values for the best α), which is the default setting of the original cPCA.

A.2 Results

Table 1 shows the completion time for each cPCA implementation. Generally, we can see that our implementations have better performance results than that of the original cPCA implementation. Also, we can see a performance improvement in the multi-threads implementation except for when $d = 1,000$. After reviewing the performance result in detail, we consider that the reason why the performance improvement cannot be seen when $d = 1,000$ is due to the computation of covariance matrices in Line 3 in Algo. 1. This is outside of the parallel computation. In particular, Line 3's time complexity is $O(nd^2)$ and becomes dominant in the computation time when d is large. When d is around 100, our cPCA implementation can complete in reasonable latency for the interactive usage (e.g., when $n = 100,000$ and $d = 100$, the completion time is 0.427s).

Table 1: Completion times of cPCA implementations.

n	d	the original cPCA	ours (single thread)	ours (multi-threads)
100	10	0.0380s	0.000538s	0.000328s
100	100	0.218s	0.0284s	0.00817s
100	1,000	25.5s	9.71s	10.1s
1,000	10	0.0450s	0.00113s	0.000490s
1,000	100	0.238s	0.0329s	0.0103s
1,000	1,000	28.5s	12.1s	14.9s
10,000	10	0.119s	0.00693s	0.00248s
10,000	100	0.479s	0.0970s	0.0362s
10,000	1,000	32.6s	13.9s	15.9s
100,000	10	1.03s	0.0839s	0.0503s
100,000	100	3.07s	0.972s	0.427s
100,000	1,000	56.9s	32.6s	26.4s

[1]The original cPCA implementation, `https://github.com/abidlabs/contrastive`, accessed: 2019-3-6

[2]The source code is available from https://takanori-fujiwara.github.io/s/dr-cl/

Algorithm 1 Our usage of cPCA with automatic selection of α

Inputs: datasets of the target cluster and the others $K = \{\mathbf{k}_i\}_{i=1}^{t}$, $R = \{\mathbf{r}_i\}_{i=1}^{u}$; list of possible $\{\alpha_i\}_{i=1}^{q}$; a threshold ratio of the variance γ.

Obtain the concatenated dataset $E = K \sqcup R = \{\mathbf{e}_i\}_{i=1}^{s}$
Apply centering to E and R
Calculate the empirical covariance matrices $\mathbf{C_E}$ and $\mathbf{C_R}$ from E and R
Perform EVD on $\mathbf{C} = \mathbf{C_E} - \alpha_1 \mathbf{C_R}$
Obtain 1D DR results K' and R' by projecting K and R with the top eigenvector of $\mathbf{C}$ (the first cPC).
Calculate the distance $D(\alpha_1)$ and variance $V(\alpha_1)$ with K' and R'
best_$\alpha = \alpha_1$, best_$D = D(\alpha_1)$, $V_{\alpha_1} = V(\alpha_1)$, best_$K' = K'$, best_$R' = R'$
for all $i = 2, 3, \cdots, q$ **do in parallel**
    Perform EVD on $\mathbf{C} = \mathbf{C_E} - \alpha_i \mathbf{C_R}$
    Obtain 1D DR results K' and R' by projecting K and R with the first cPC
    Calculate the distance $D(\alpha_i)$ and variance $V(\alpha_i)$ with K' and R'
    if $D(\alpha_i) >$ best_D and $V(\alpha_i) \geq \gamma V_{\alpha_1}$ **then**
        best_$\alpha = \alpha_i$, best_$D = D(\alpha_i)$, best_$K' = K'$, best_$R' = R'$
return best_α, best_K', best_R'

B ccPCA Results with Different α Values

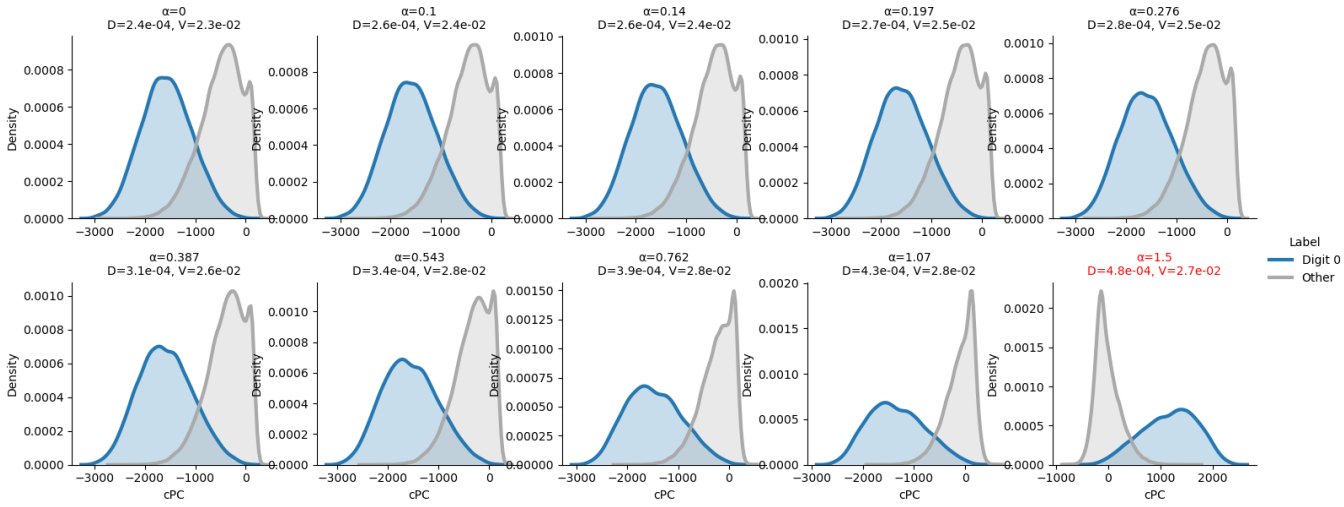
To supplement Sect. 4.2.3, we provide additional examples of several datasets (the MNIST, Wine Recognition, Tennis Major Tournament Match Statistics, Nutrients, Communities and Crime datasets). In particular, we show 1D dimensionality reduction (DR) results with the first cPC obtained with different α values. DR results show the quality of the ccPCA results in terms of the separation between the target cluster and the others and the preservation of the target cluster's variance. Also, we highlight the best α selected by our automatic selection algorithm. Additionally, for the MNIST dataset, we provide visualizations of the features relative contributions (FCs) to the first cPC. We can then visually see where our selection algorithm decides which important features to contrast the target digit with the other digits.

The data we have analyzed are all available at `https://vis-sub.bitbucket.io/dr-cl/`, which includes all the features, 2D positions of the projection with t-SNE, and cluster labels assigned by DBSCAN. For each example, we show the stated results with ten different α values. We use density plots to depict the 1D DR results and a 2D heatmap for the FCs of MNIST digits. Also, we indicate the best α selected by our algorithm by using a red subtitle in the corresponding figure.

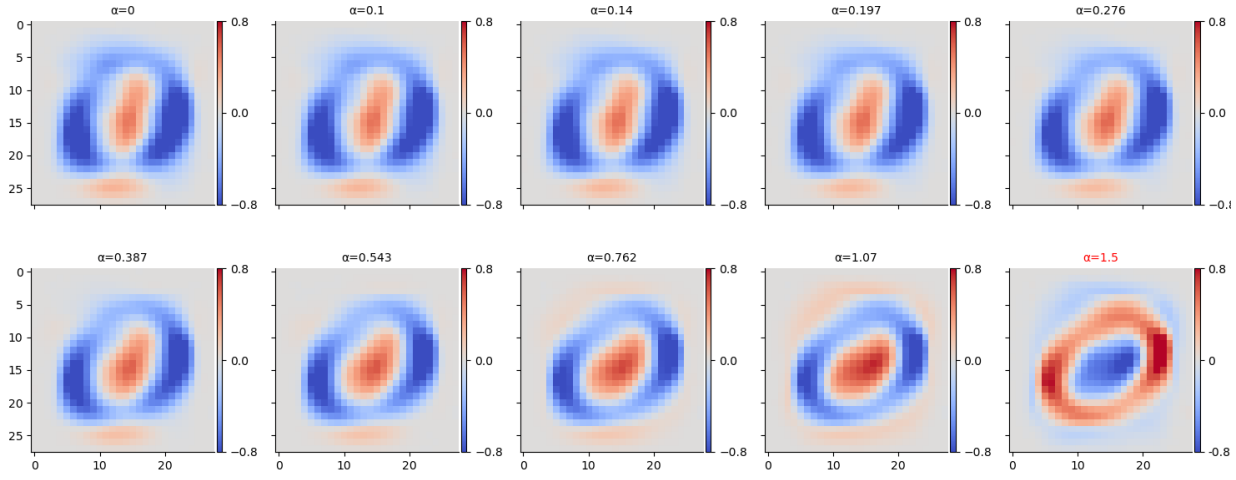
B.1 MNIST Dataset

From Fig. B.1 to Fig. B.10, we show the results of the MNIST dataset. We have generated 10 candidate α values that are logarithmically spaced in a range between 0 and 1.5. Also, we set $\gamma = 0.5$ (a ratio controlling the threshold of the variance V).

Generally, we can see that while ccPCA does not find the features (i.e., pixels) contrasting the target digit with the others when α is close to 0, ccPCA gradually starts to find such features as α increases. However, when α is larger than an optimal value, ccPCA selects less commonly used pixels (e.g., Fig. B.10b with $\alpha = 1.5$). Our automatic selection algorithm chooses a well-balanced α value to contrast the target digit.

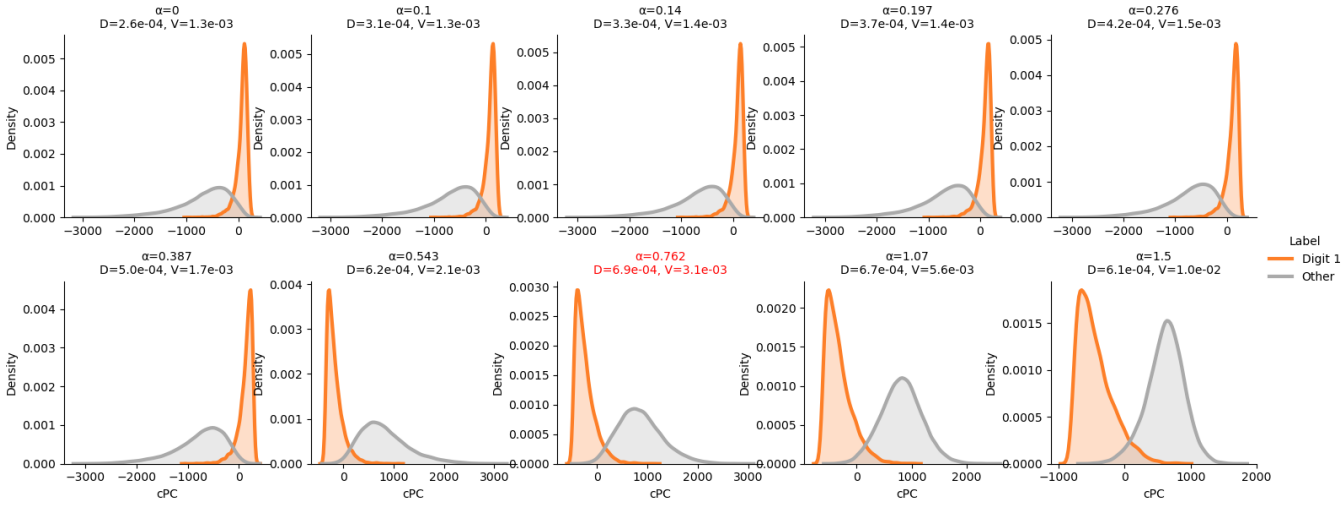


(a) 1D DR result

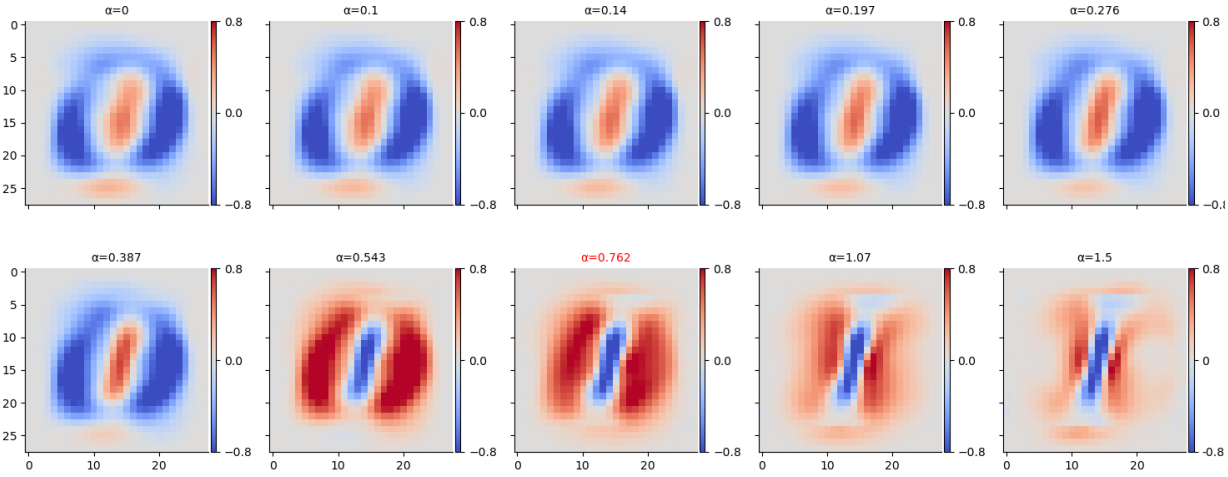


(b) Feature contributions

Figure B.1: MNIST Digit 0

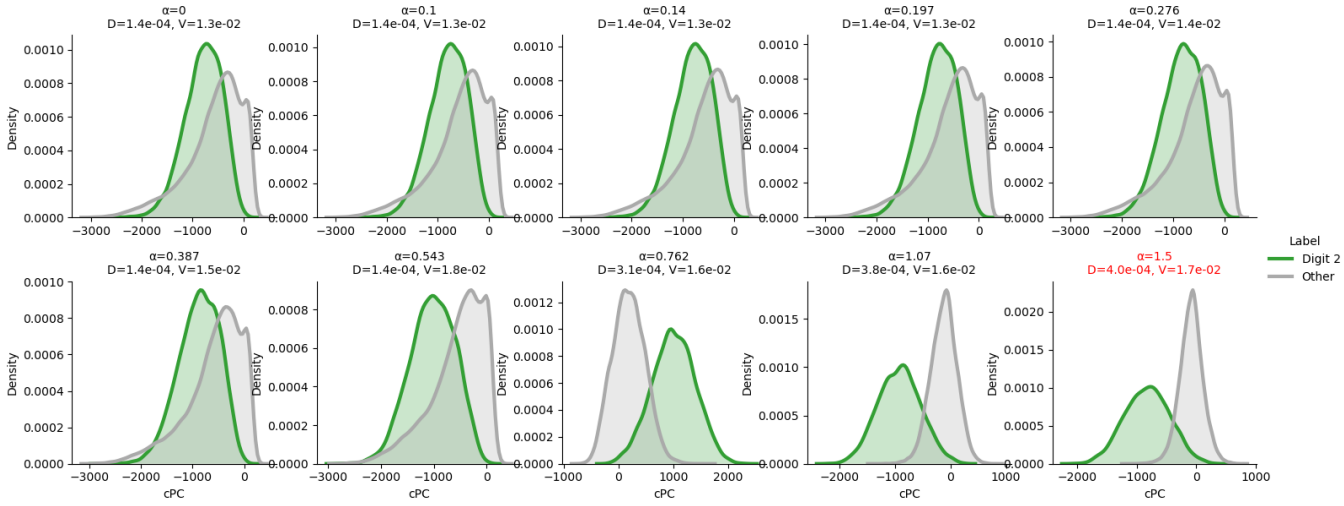


(a) 1D DR result

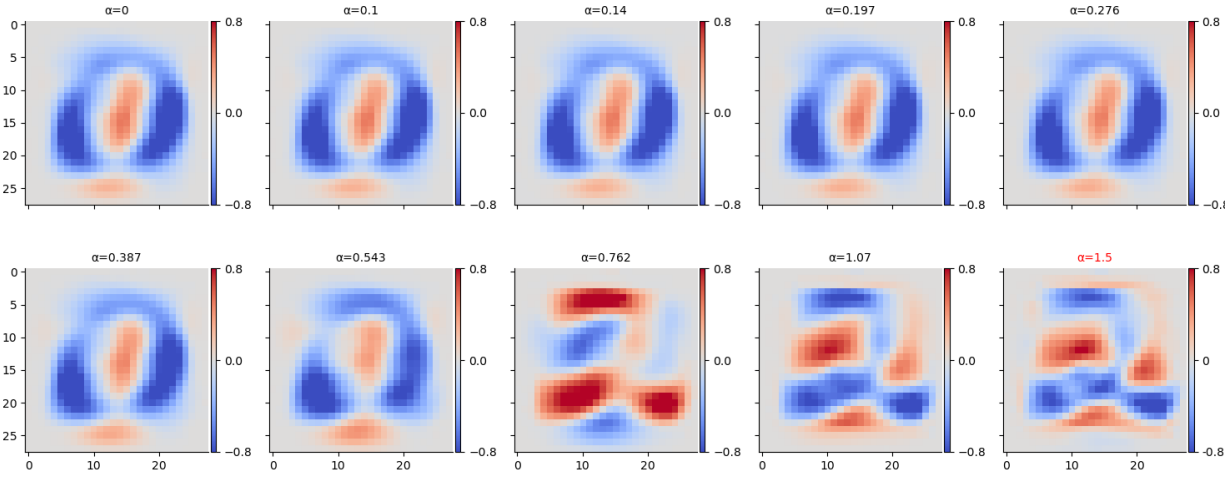


(b) Feature contributions

Figure B.2: MNIST Digit 1

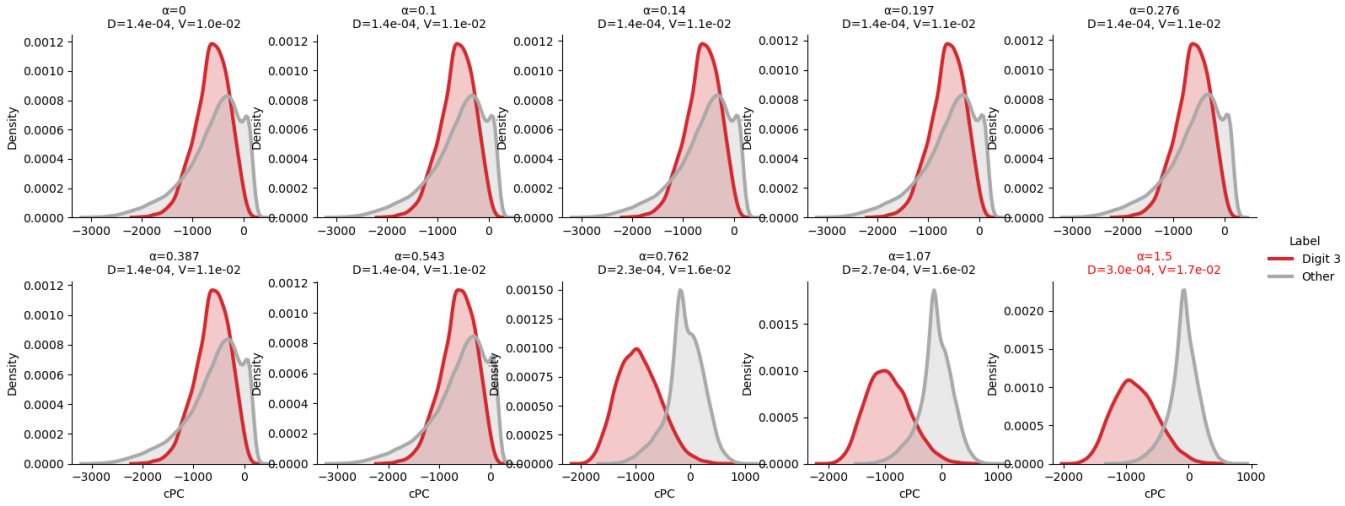


(a) 1D DR result

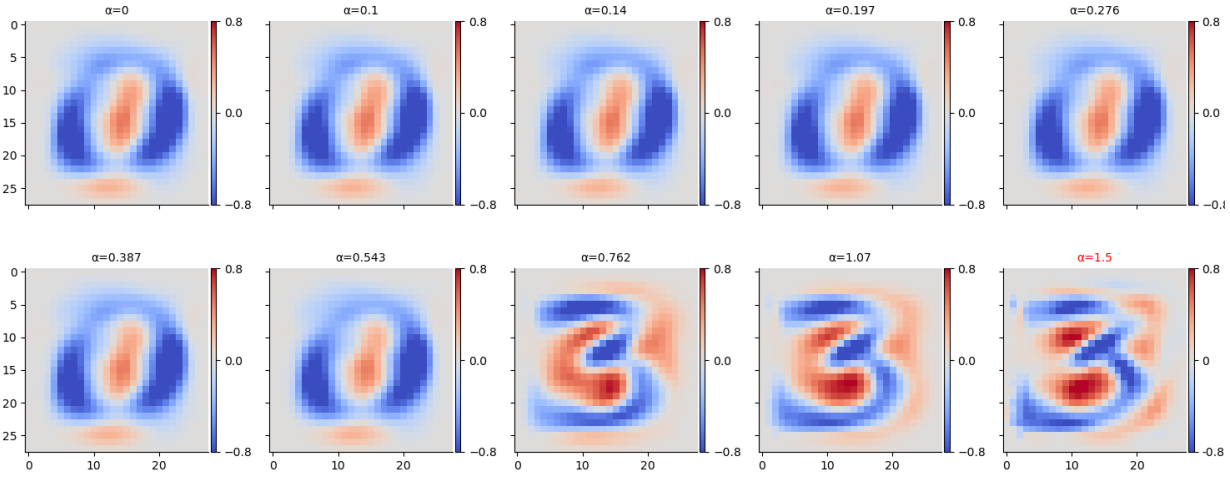


(b) Feature contributions

Figure B.3: MNIST Digit 2

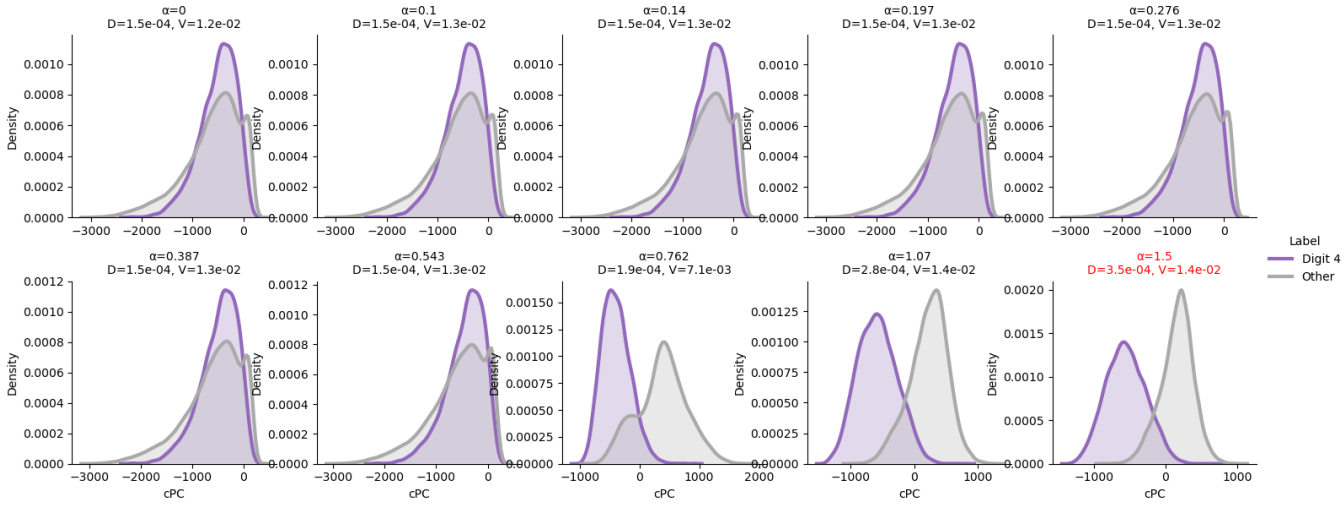


(a) 1D DR result

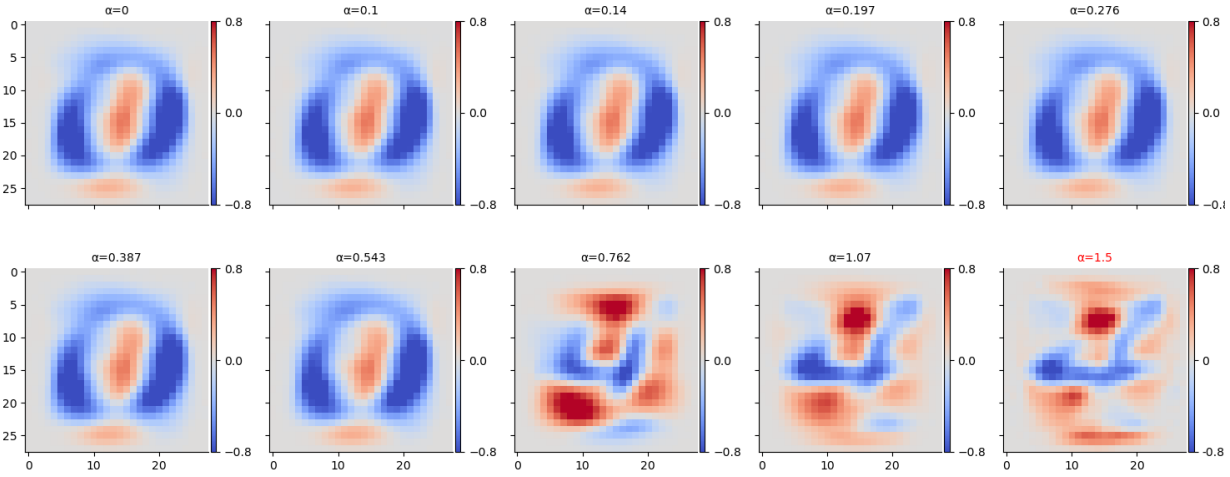


(b) Feature contributions

Figure B.4: MNIST Digit 3

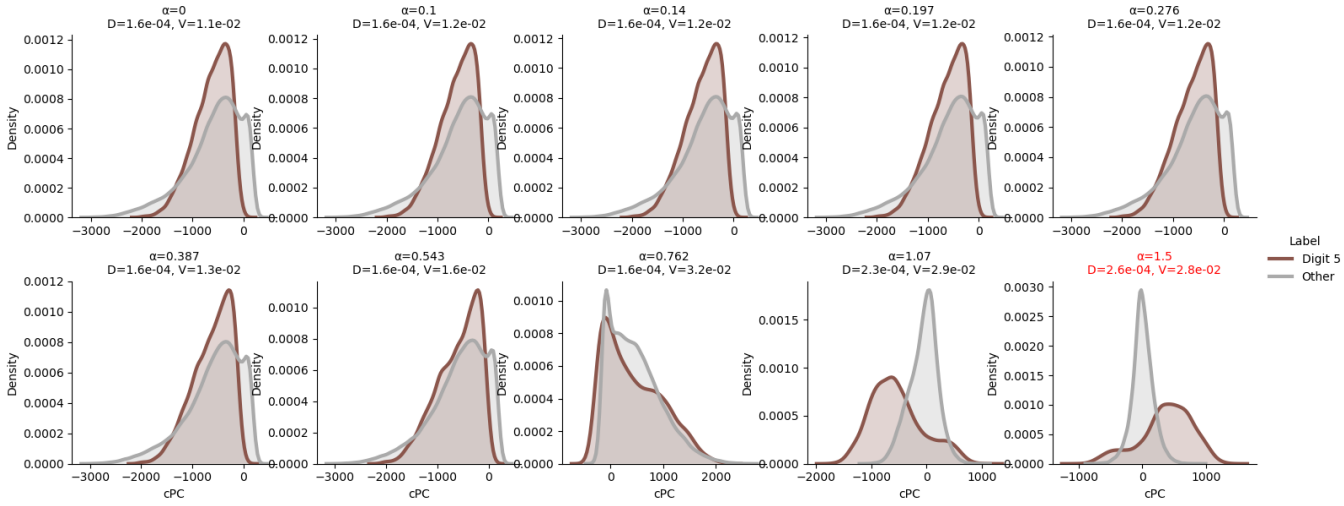


(a) 1D DR result

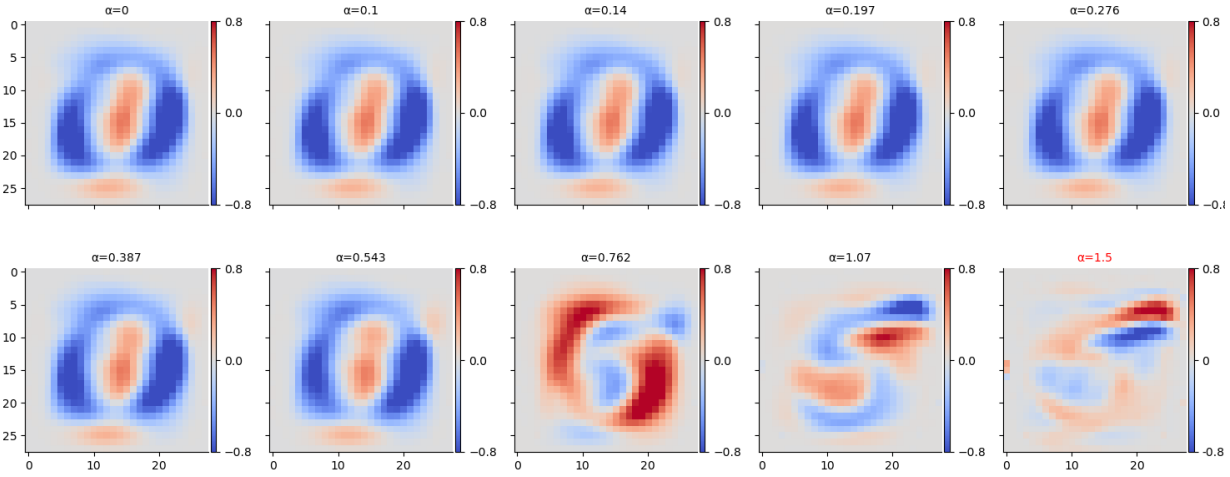


(b) Feature contributions

Figure B.5: MNIST Digit 4

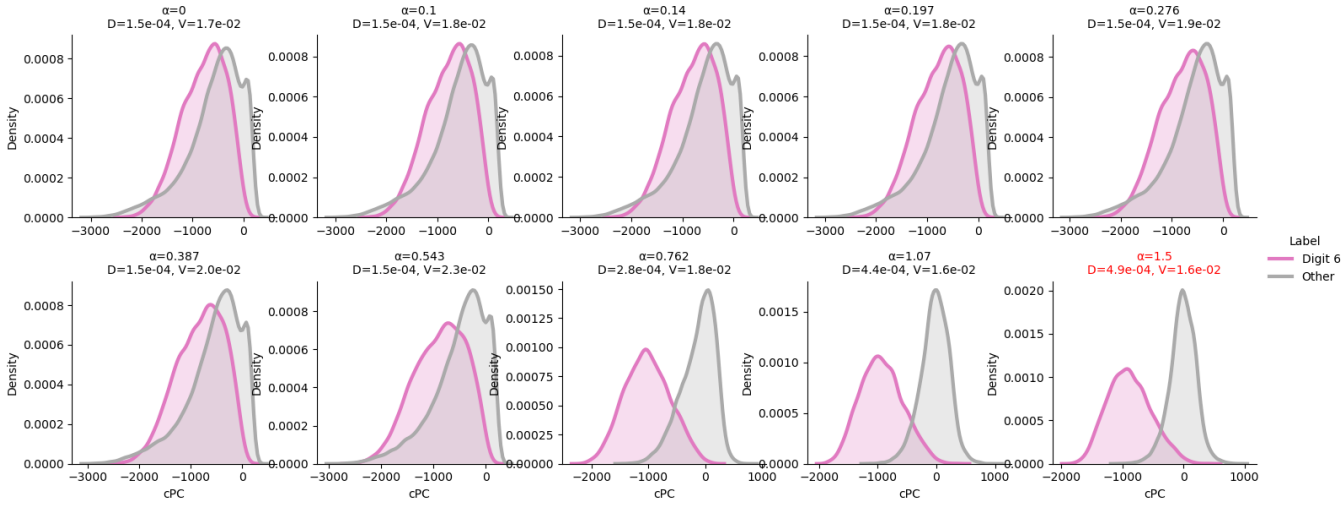


(a) 1D DR result

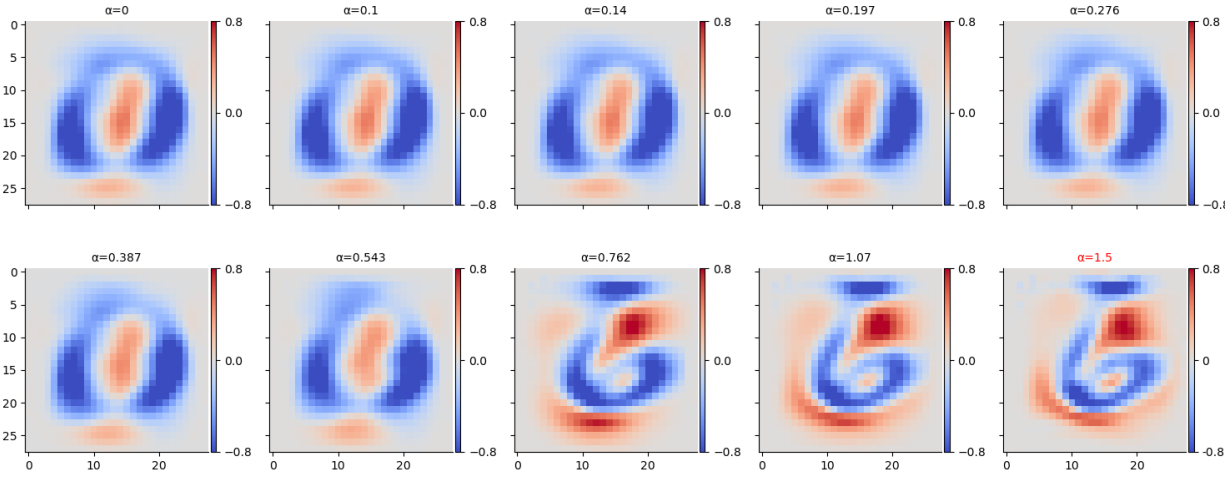


(b) Feature contributions

Figure B.6: MNIST Digit 5

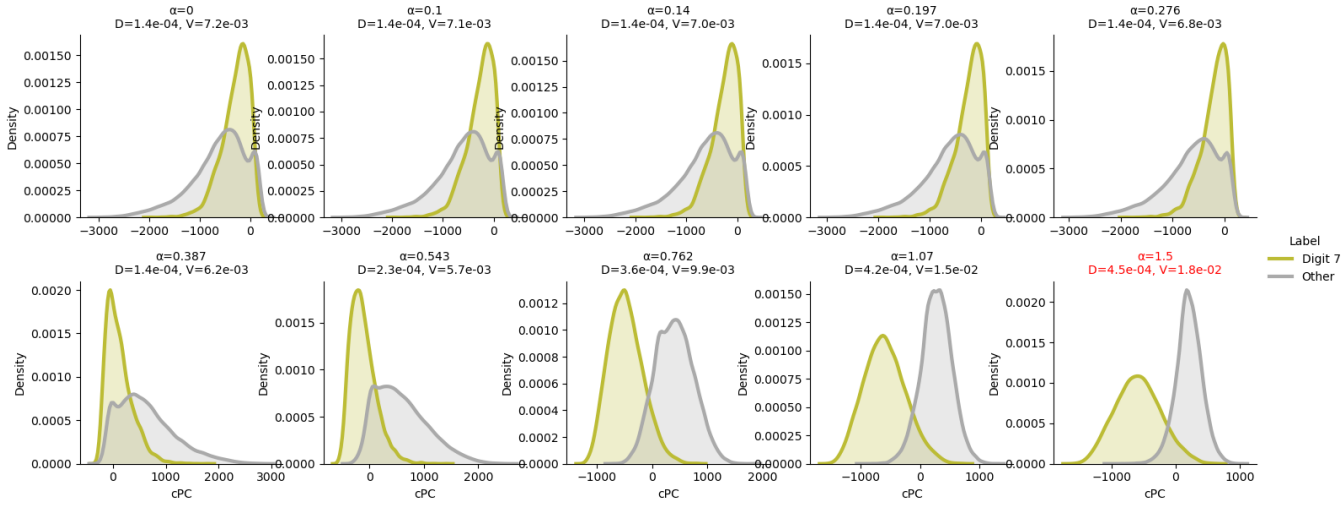


(a) 1D DR result

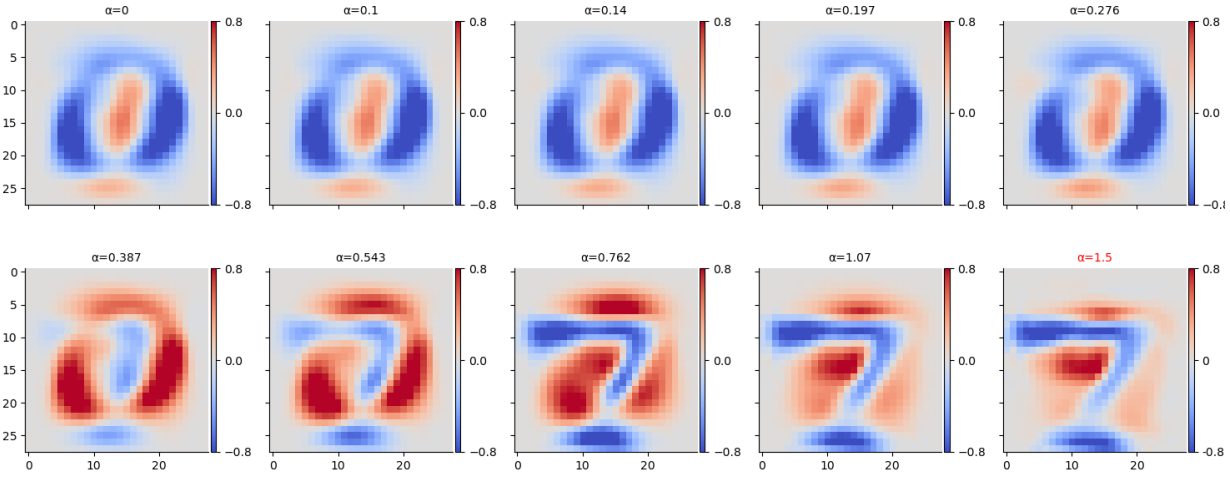


(b) Feature contributions

Figure B.7: MNIST Digit 6

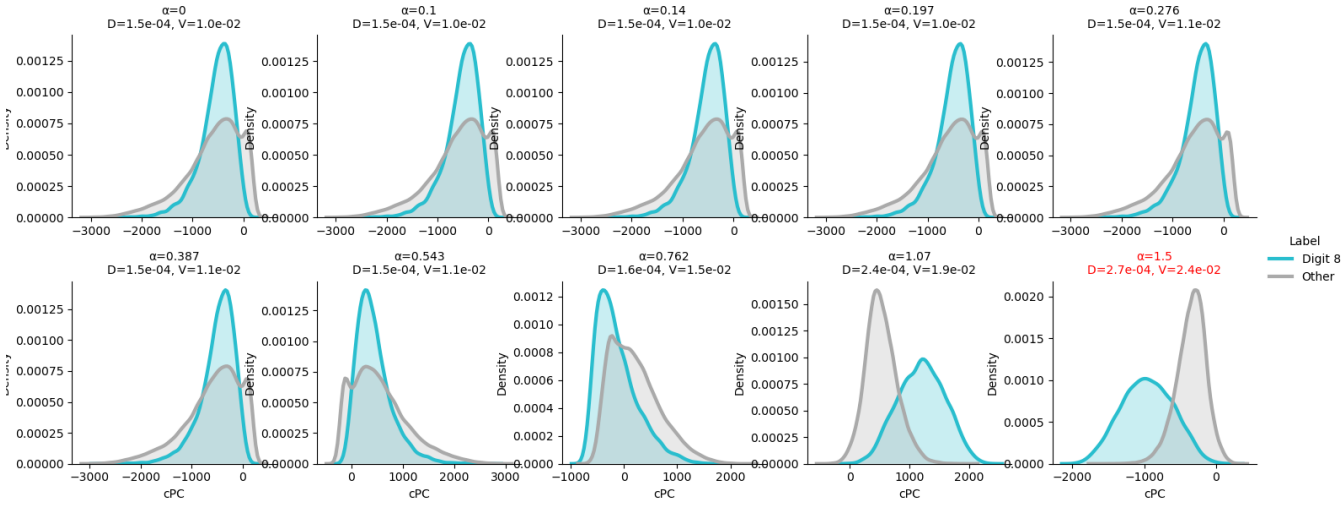


(a) 1D DR result

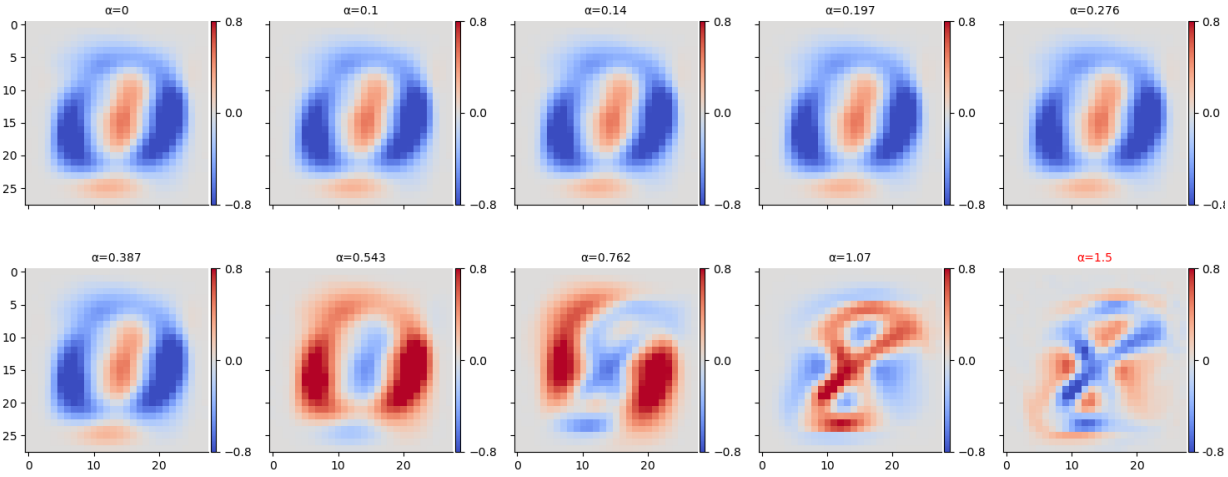


(b) Feature contributions

Figure B.8: MNIST Digit 7

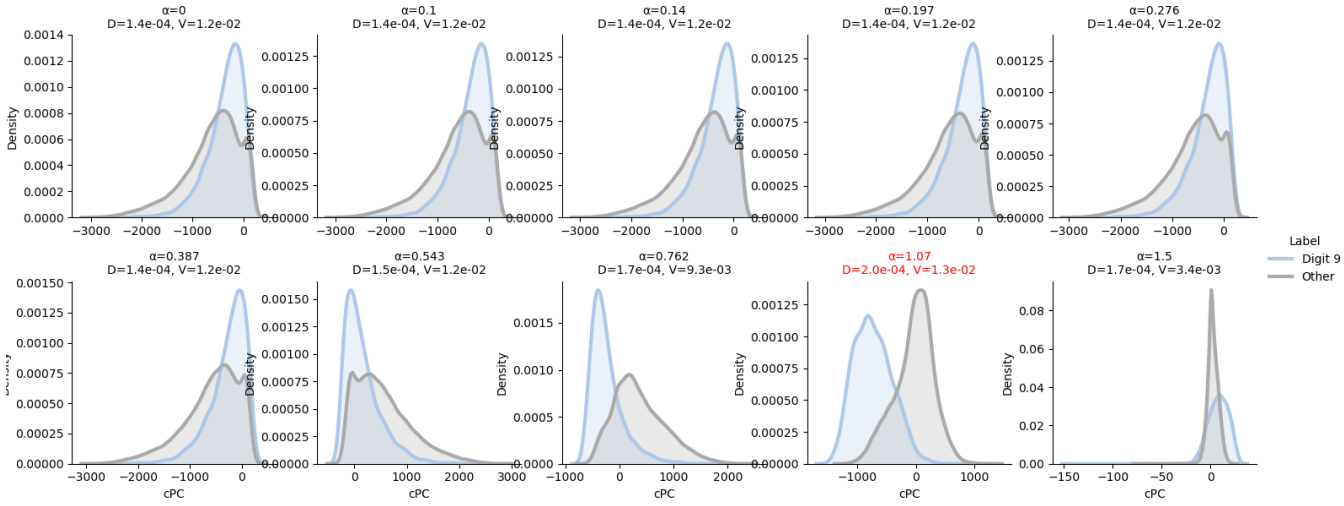


(a) 1D DR result

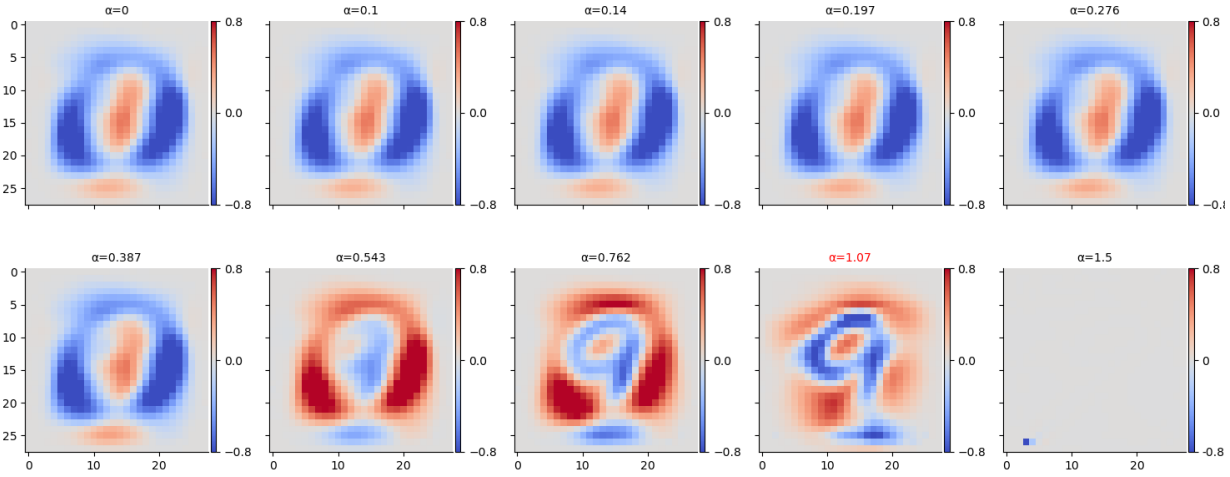


(b) Feature contributions

Figure B.9: MNIST Digit 8



(a) 1D DR result



(b) Feature contributions

Figure B.10: MNIST Digit 9

B.2 Wine Recognition Dataset

From Fig. B.11 to Fig. B.13, we show the results of the Wine Recognition dataset. We have generated 10 candidate α values that are logarithmically spaced in a range between 0 and 10. Also, we set $\gamma = 0.5$.

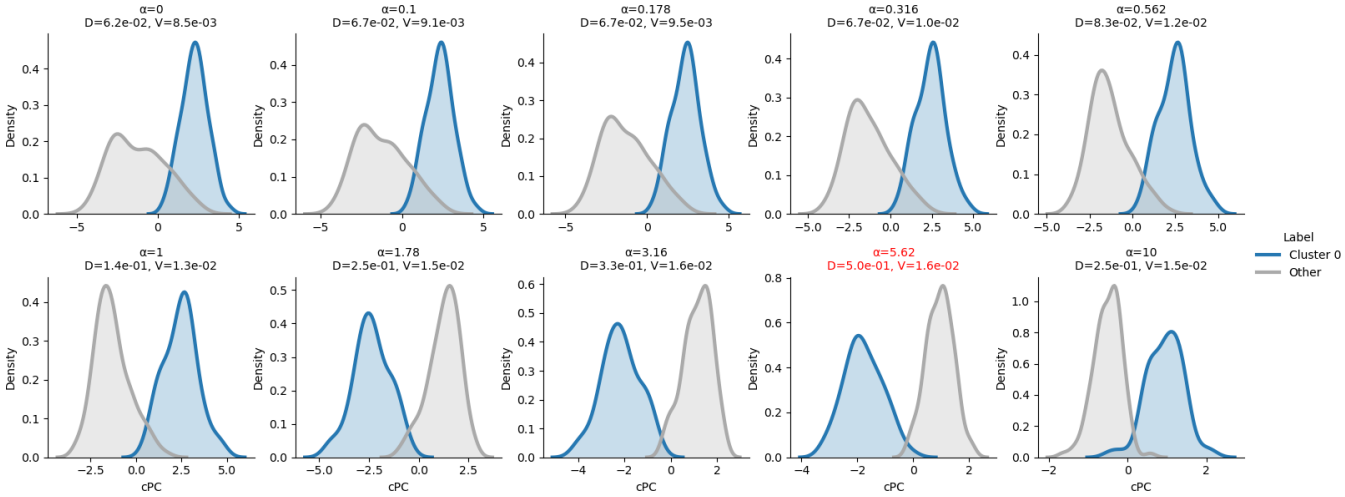


Figure B.11: Wine Recognition: DR result of Cluster 0

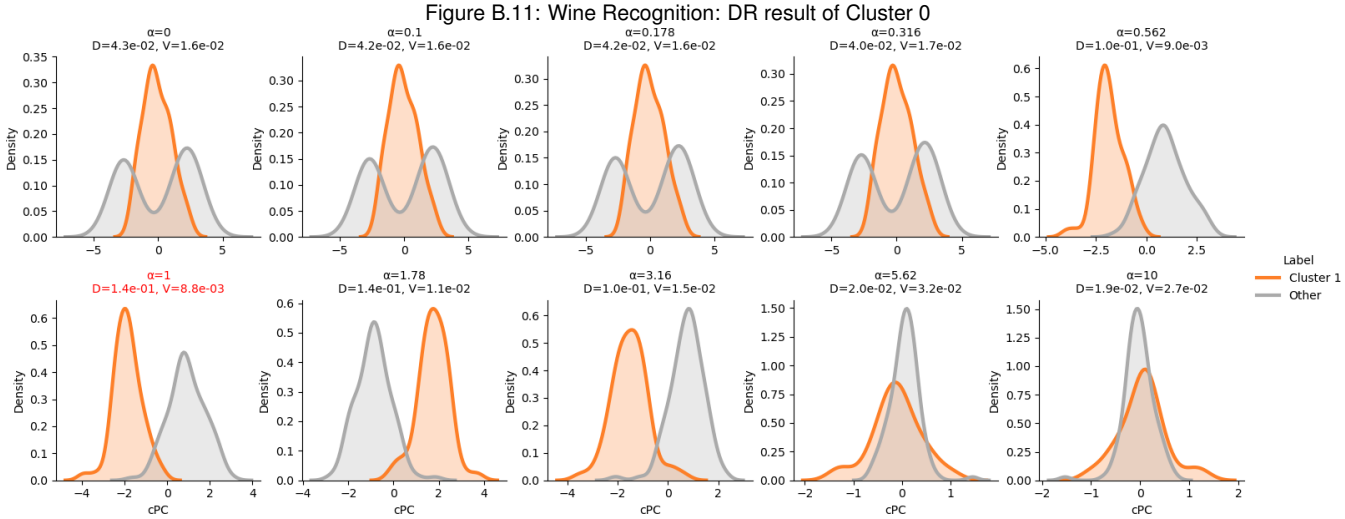


Figure B.12: Wine Recognition: DR result of Cluster 1

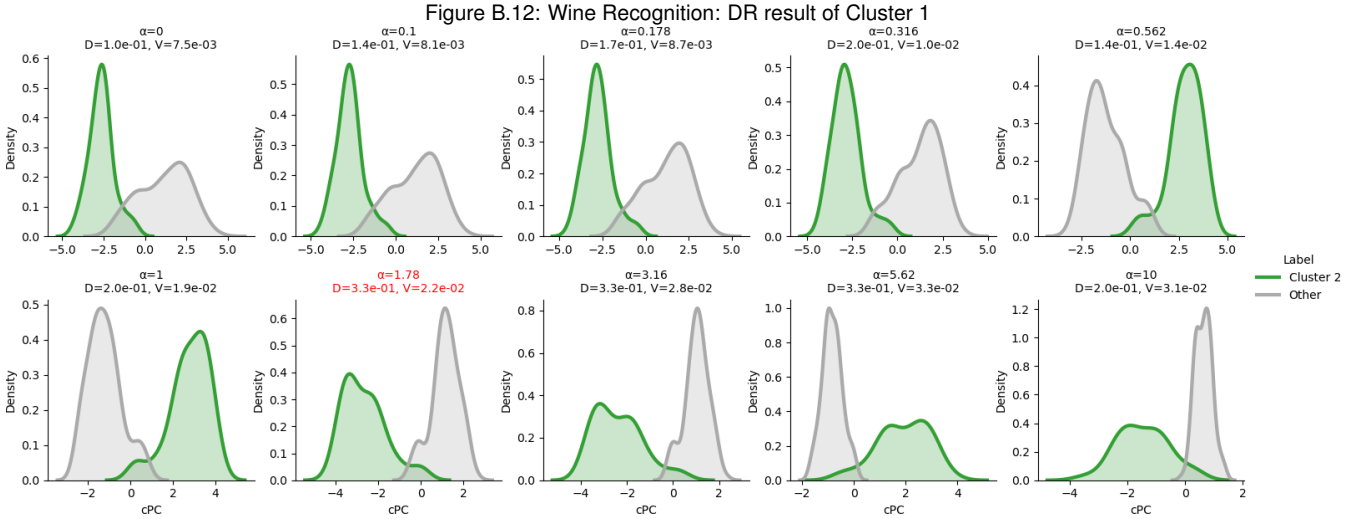


Figure B.13: Wine Recognition: DR result of Cluster 2

B.3 Tennis Major Tournament Match Statistics Dataset

From Fig. B.14 to Fig. B.18, we show the results of the Tennis Major Tournament Match Statistics dataset. We have generated 10 candidate α values that are logarithmically spaced in a range between 0 and 10. Also, we set $\gamma = 0.5$.

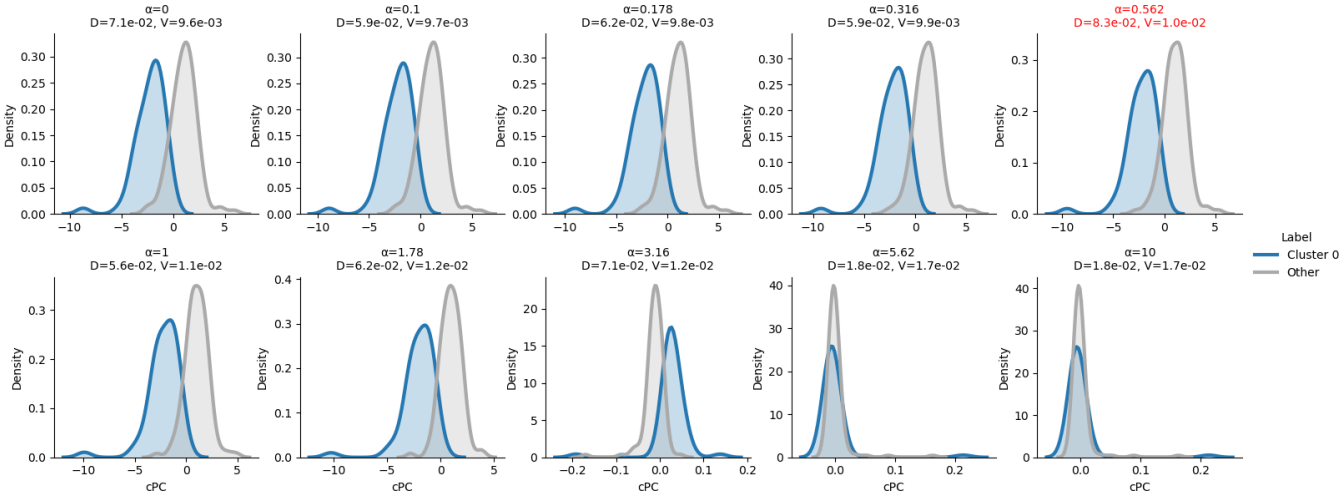


Figure B.14: Tennis Major Tournament Match Statistics: DR result of Cluster 0

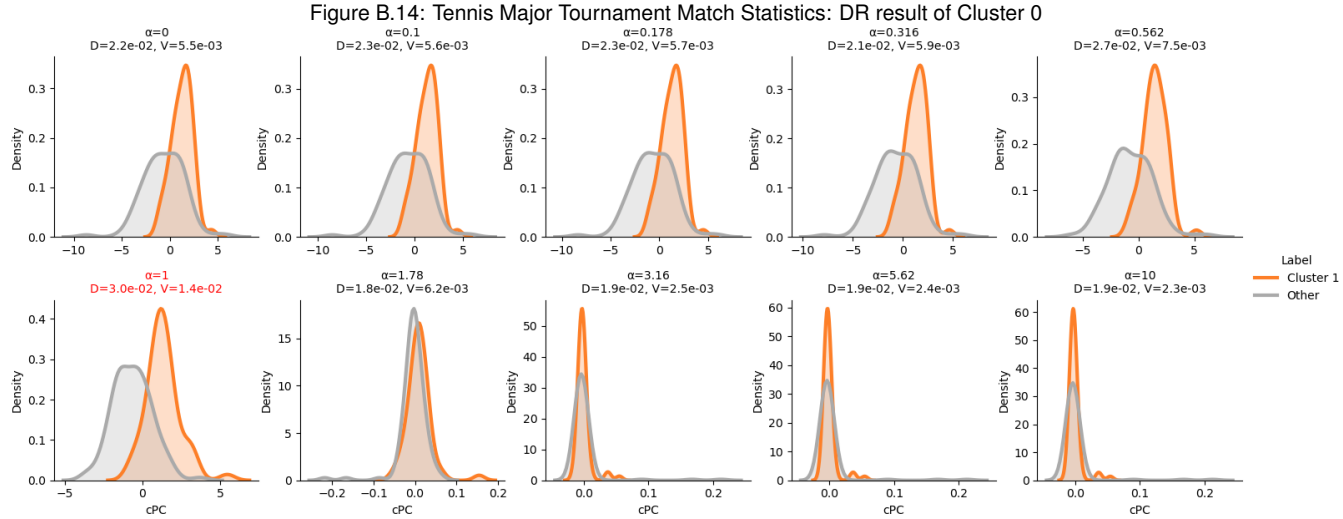


Figure B.15: Tennis Major Tournament Match Statistics: DR result of Cluster 1

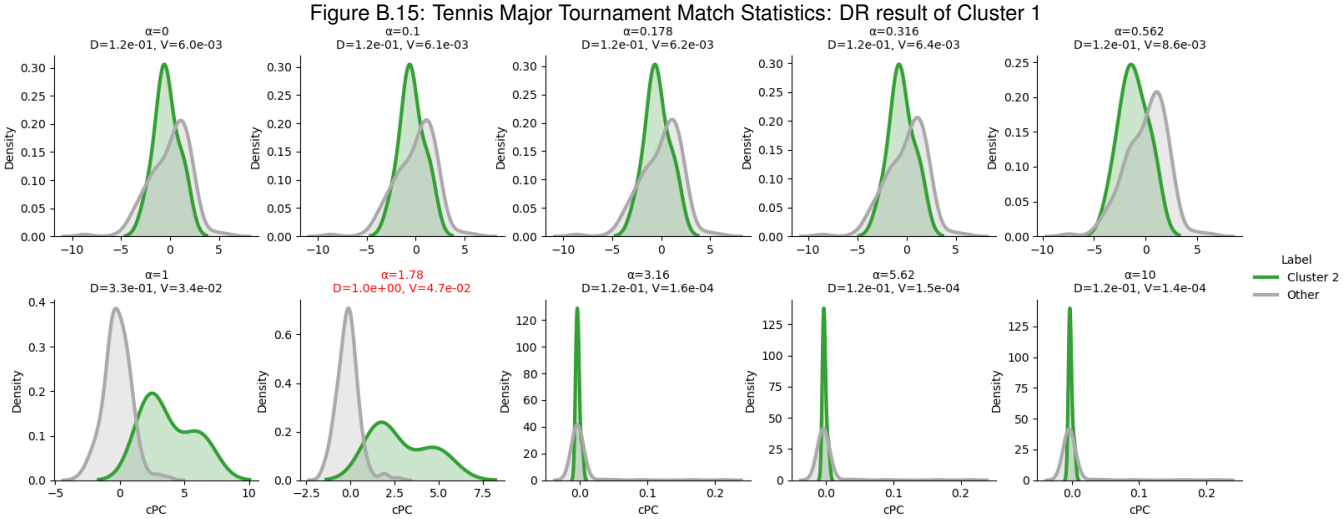


Figure B.16: Tennis Major Tournament Match Statistics: DR result of Cluster 2

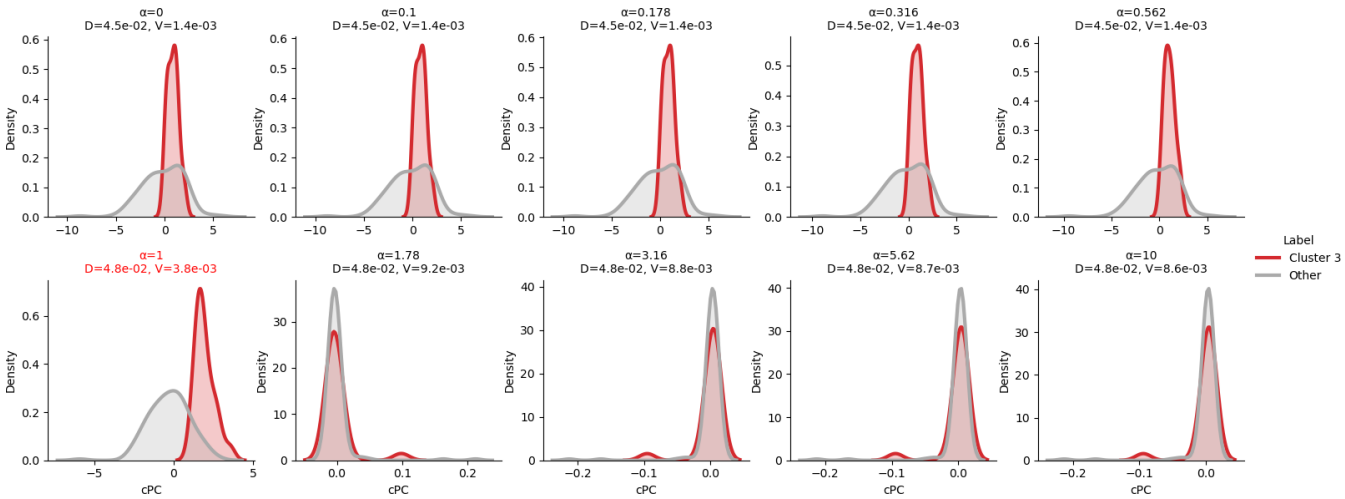


Figure B.17: Tennis Major Tournament Match Statistics: DR result of Cluster 3

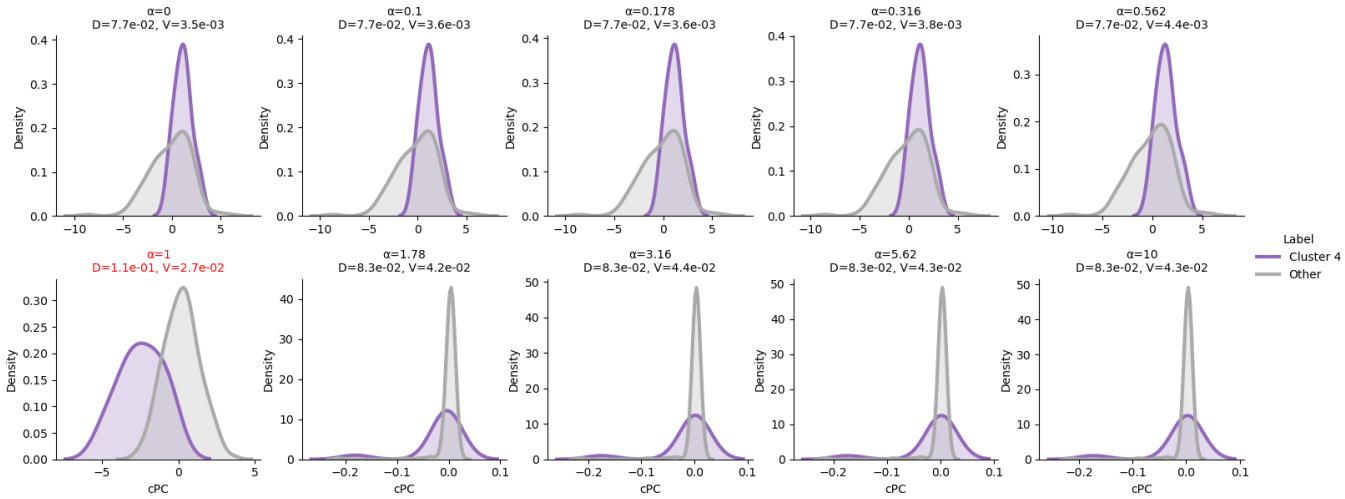


Figure B.18: Tennis Major Tournament Match Statistics: DR result of Cluster 4

B.4 Nutrients Dataset

From Fig. B.19 to Fig. B.27, we show the results of the Nutrients dataset. We have generated 10 candidate α values that are logarithmically spaced in a range between 0 and 10. Also, we set $\gamma = 0.5$. When compared to the other datasets, our automatic selection tends to select $\alpha = 0$. As described in Sect. 6.2, because some specific features ('calories' and 'fat') are dominant to contrast each target cluster, we consider that the classical PCA (corresponds to ccPCA with $\alpha = 0$) can produce well-contrasted results for some of the clusters.

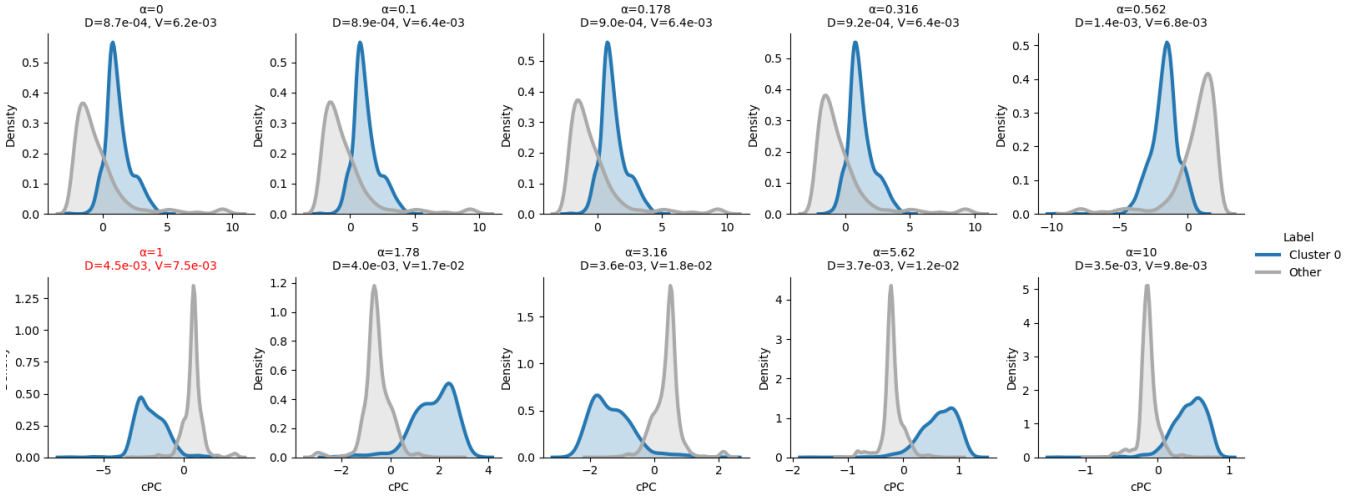


Figure B.19: Nutrients: DR result of Cluster 0

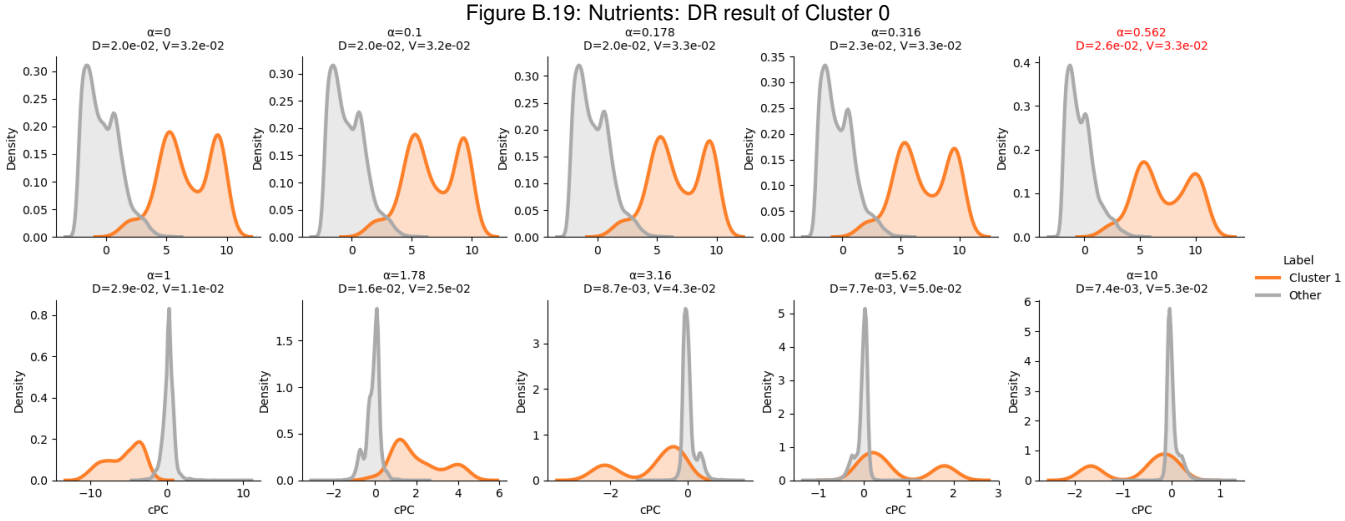


Figure B.20: Nutrients: DR result of Cluster 1

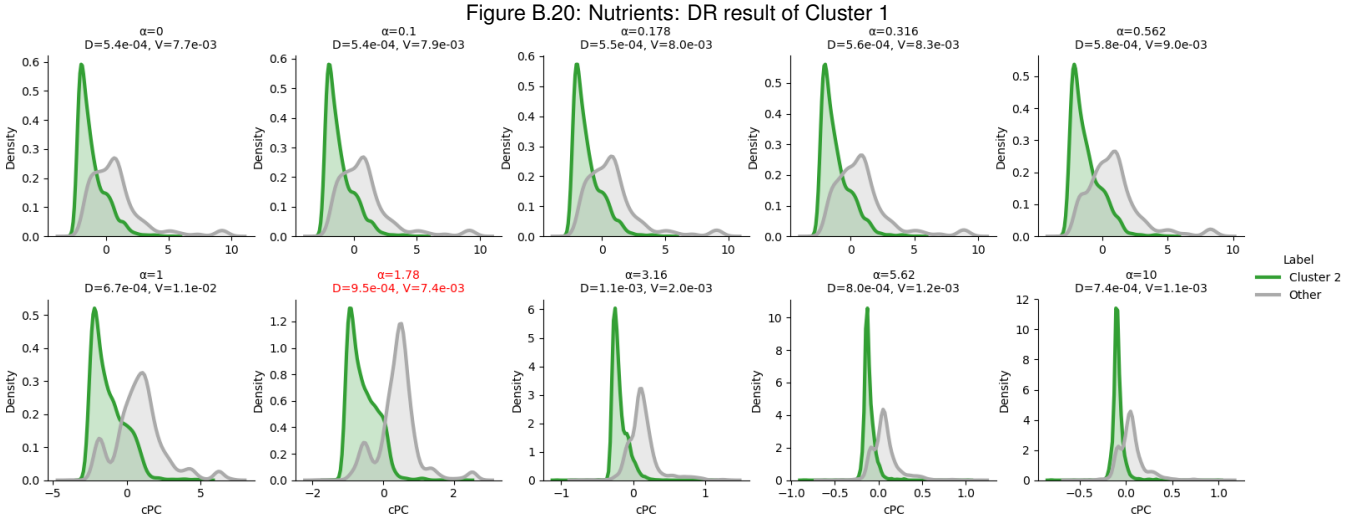


Figure B.21: Nutrients: DR result of Cluster 2

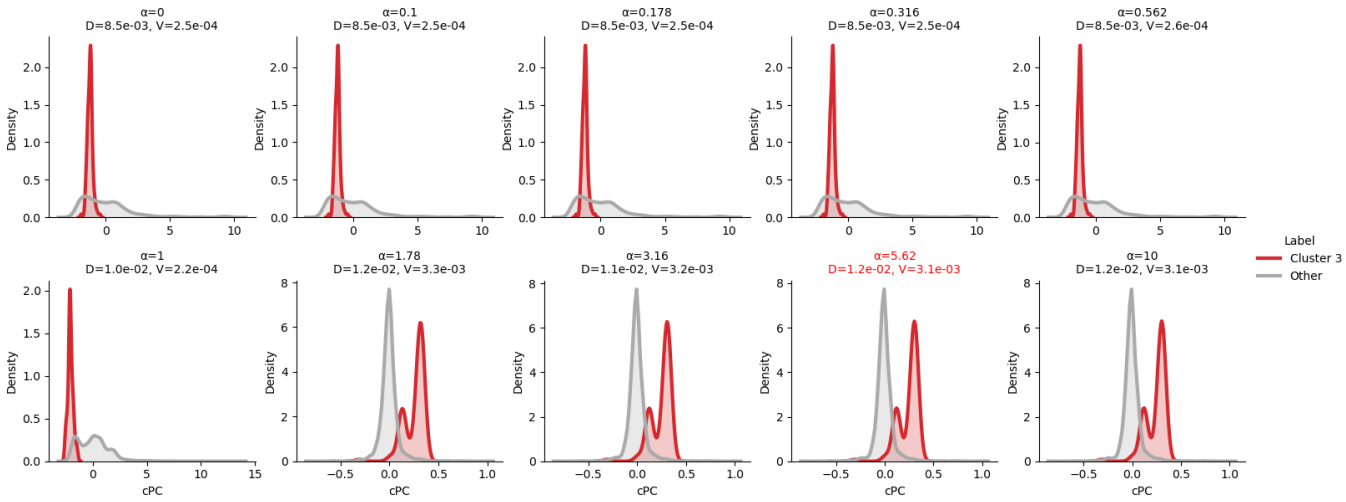


Figure B.22: Nutrients: DR result of Cluster 3

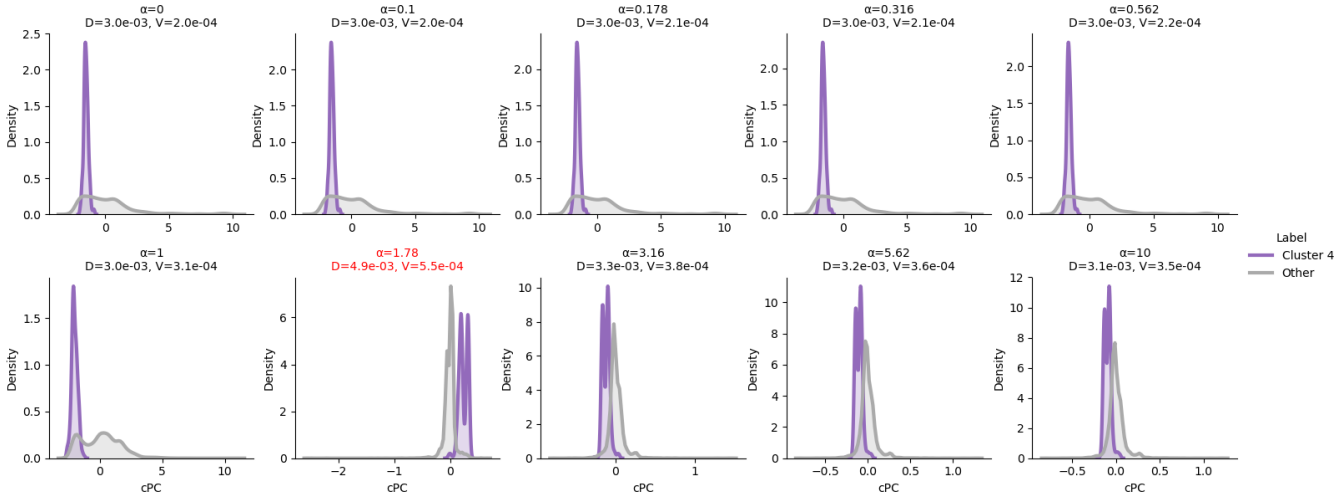


Figure B.23: Nutrients: DR result of Cluster 4

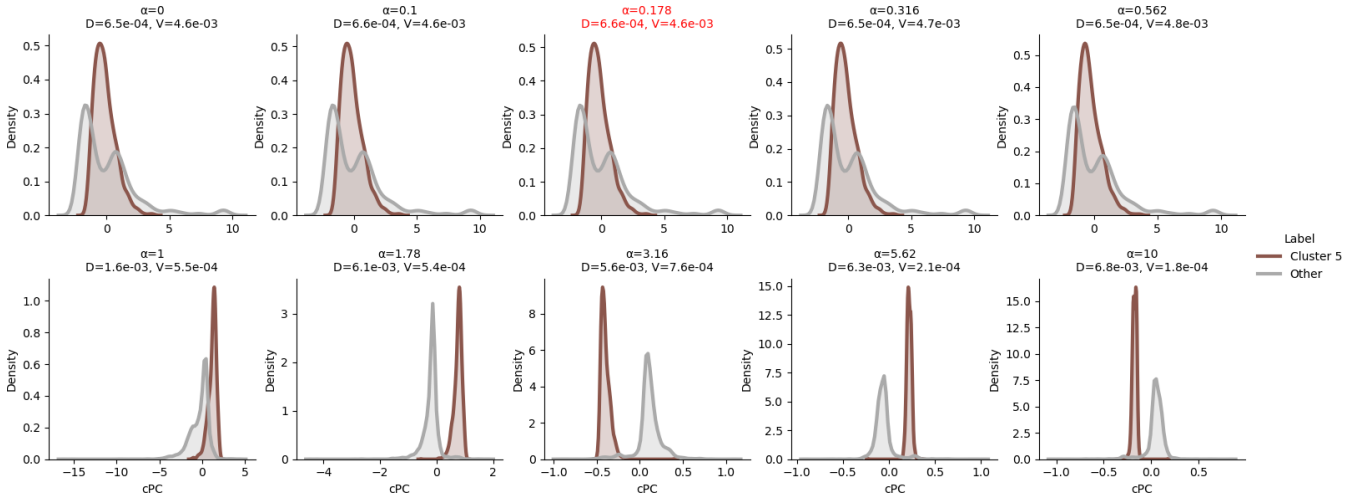


Figure B.24: Nutrients: DR result of Cluster 5

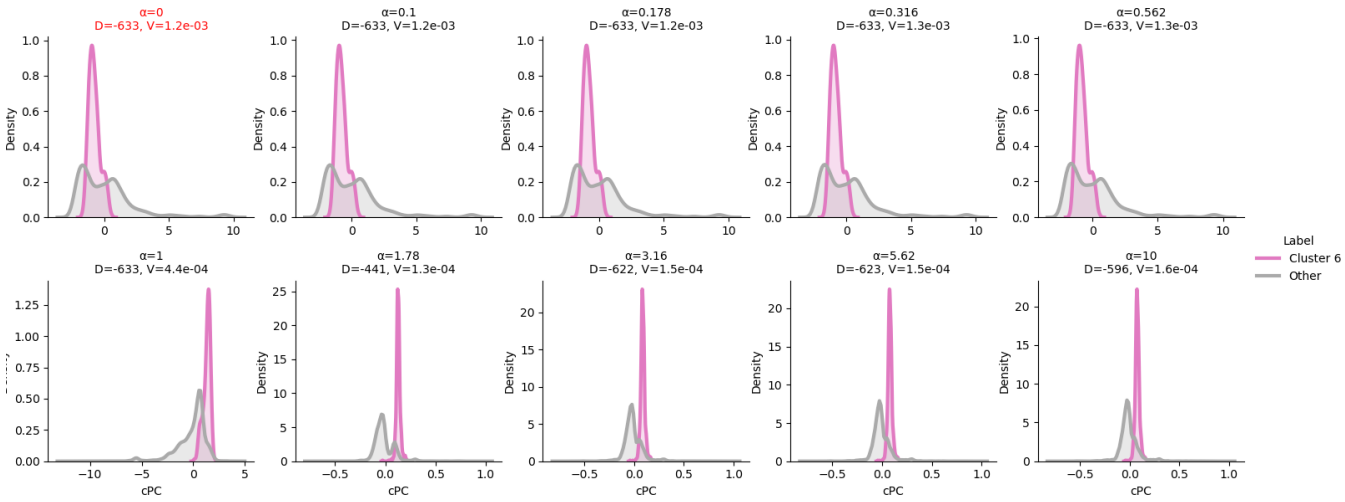


Figure B.25: Nutrients: DR result of Cluster 6

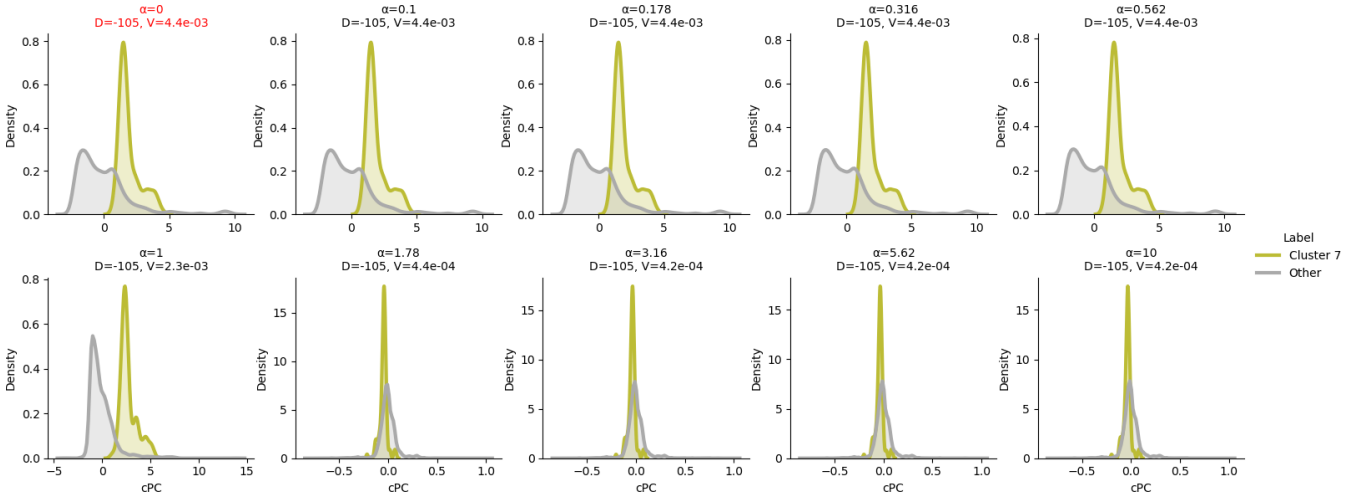


Figure B.26: Nutrients: DR result of Cluster 7

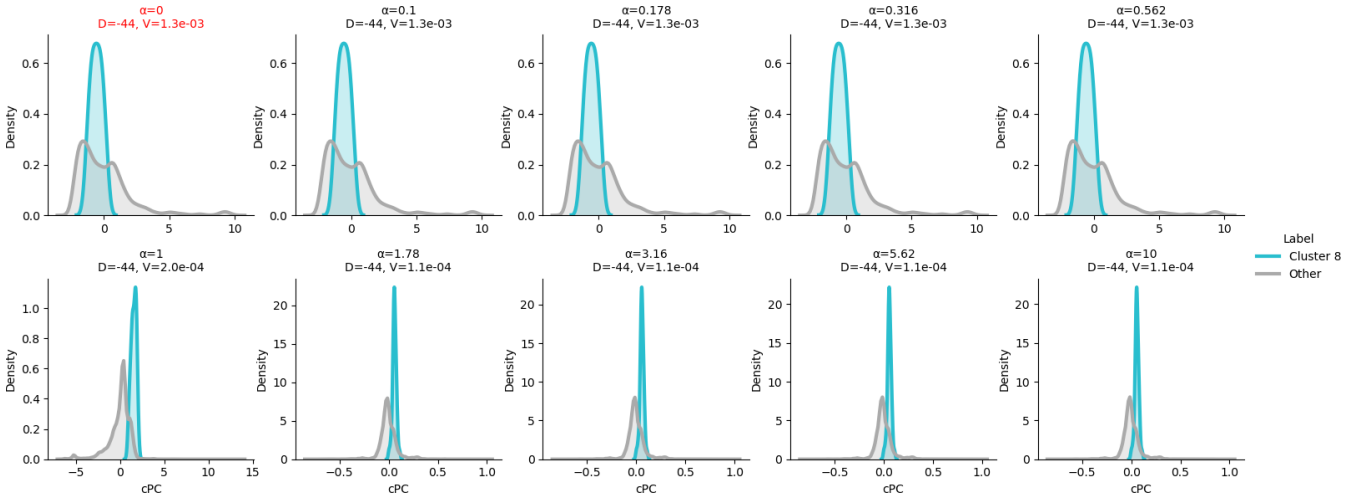


Figure B.27: Nutrients: DR result of Cluster 8

B.5 Communities and Crime Dataset

From Fig. B.28 to Fig. B.32, we show the results of the Communities and Crime dataset. We have generated 10 candidate α values that are logarithmically spaced in a range between 0 and 10. Also, we set $\gamma = 0.5$.

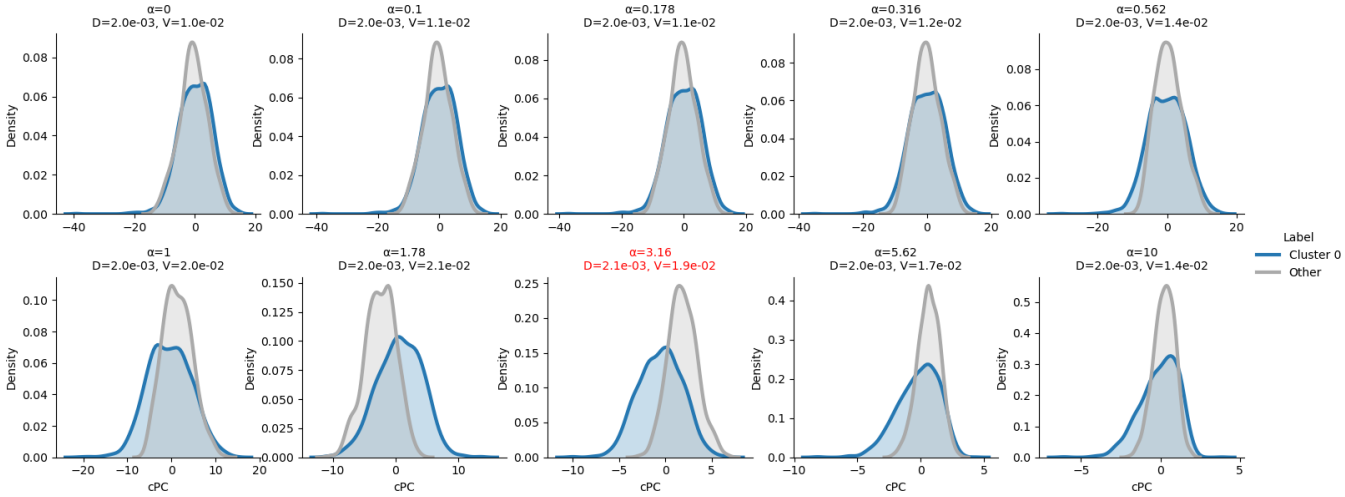


Figure B.28: Communities and Crime: DR result of Cluster 0

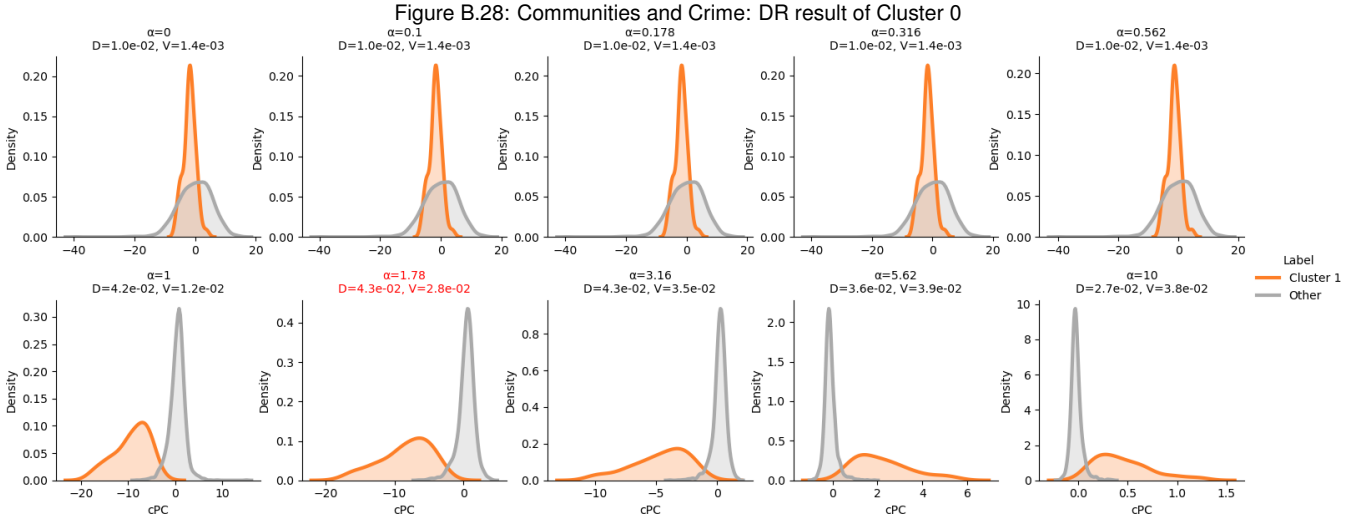


Figure B.29: Communities and Crime: DR result of Cluster 1

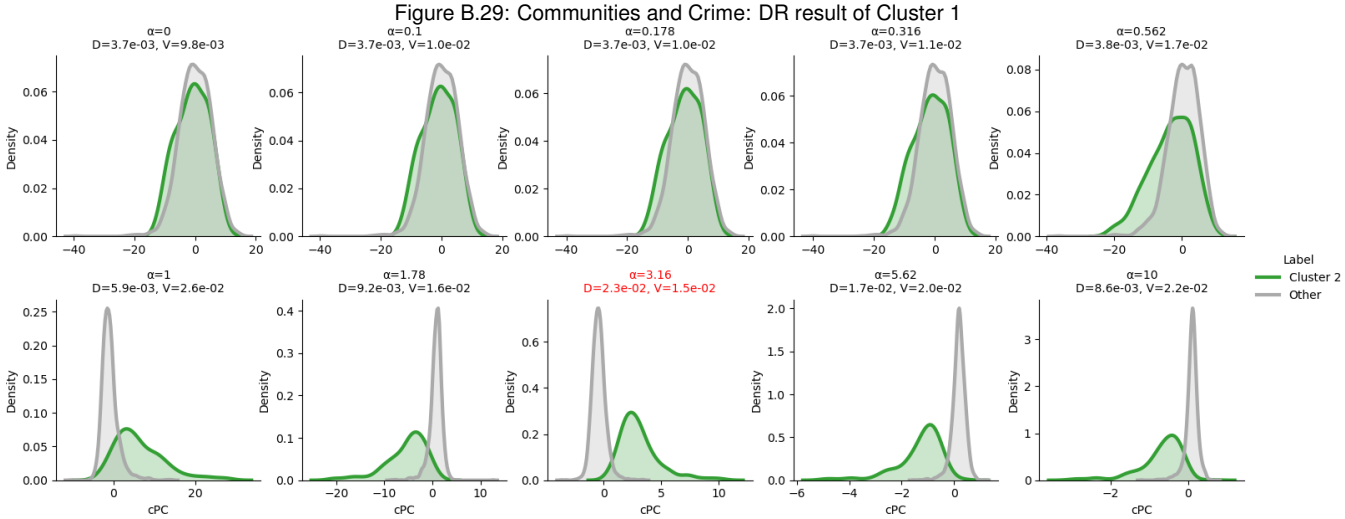


Figure B.30: Communities and Crime: DR result of Cluster 2

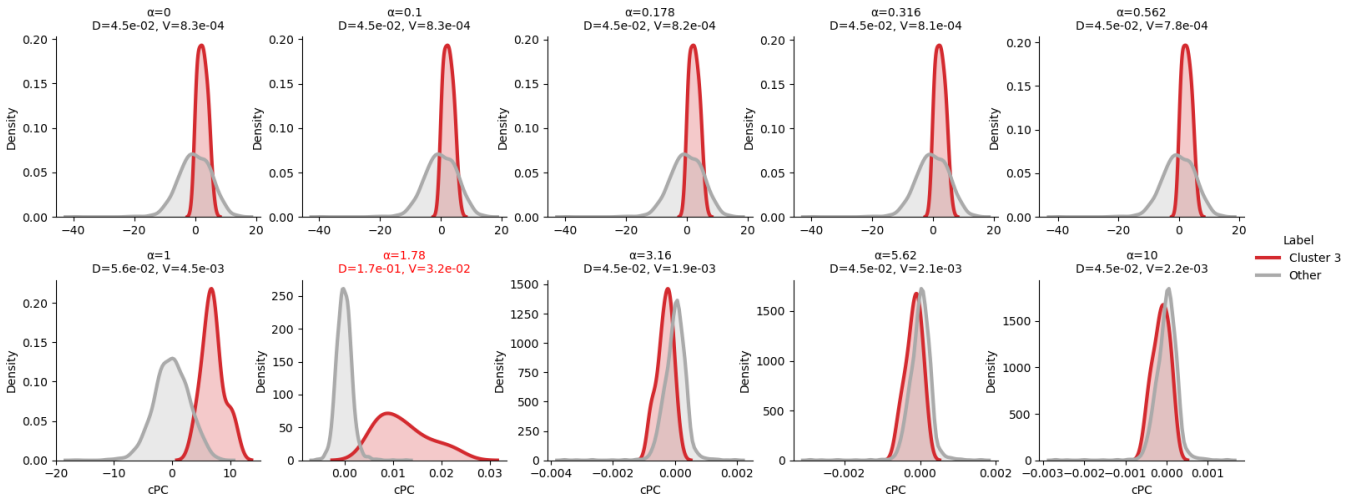


Figure B.31: Communities and Crime: DR result of Cluster 3

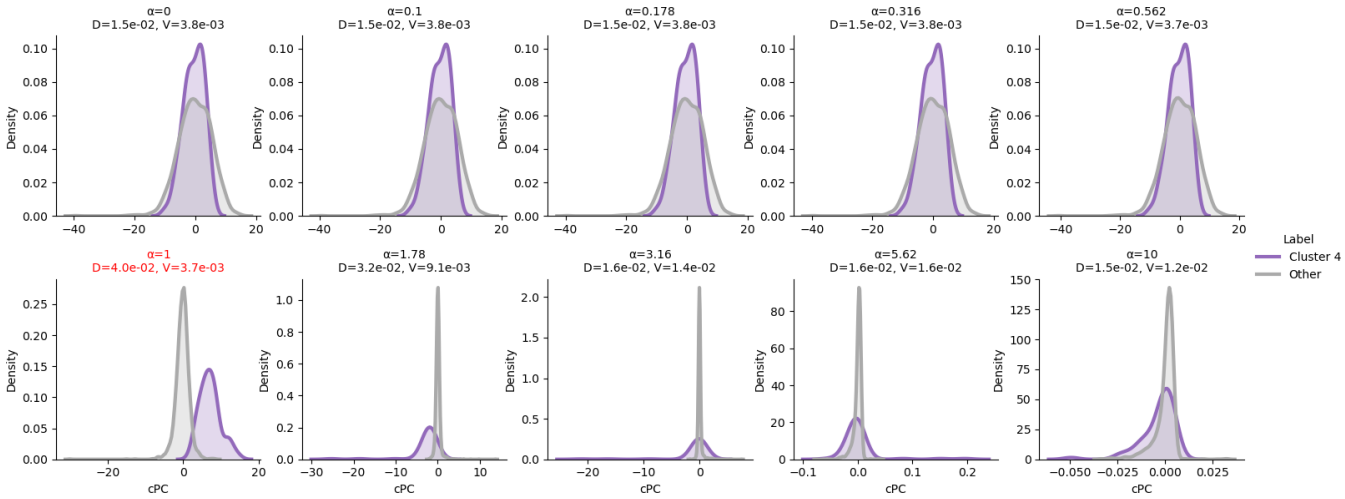


Figure B.32: Communities and Crime: DR result of Cluster 4

C Full Comparison of LDA, PCA, and ccPCA

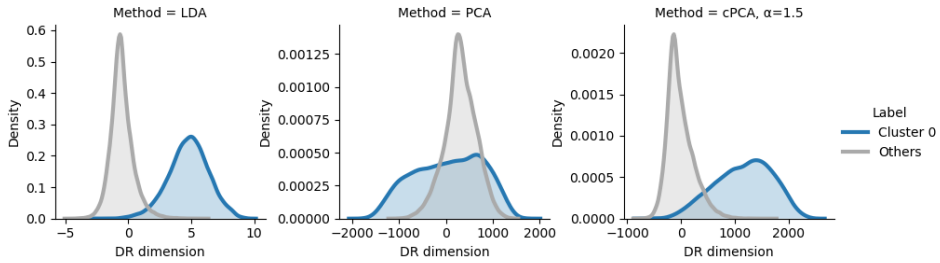
In Sect. 4.3, we show the comparison of DR methods which we can obtain the features' relative contributions to each of the first components. Here, we show more comprehensive comparison results, using the full set of MNIST, Wine Recognition, Tennis Major Tournament Match Statistics, Nutrients, and Communities and Crime datasets. In particular, we show 1D DR results using the first component obtained by LDA, PCA, or ccPCA. Also, for each result, we provide distributions of features' values which are in the top 3 contributions. With these distributions, we can see how each of the methods (i.e., LDA, PCA, or ccPCA) characterize the target cluster. Additionally, for the MNIST dataset, we provide visualizations of the features relative contributions to the first component. With these, we can visually see where each of the methods decides as the important features to contrast the target digit from the other digits.

Similar to Fig. 8 in Sect. 4.3, for LDA, we use different class labels for the target cluster and the others (e.g., Label 0 to Digit 8 and Label 1 to the other digits). For PCA, we apply it to only the target cluster. For ccPCA, we use the best α obtained from our automatic selection method with the default setting. We use density plots to depict the 1D DR results and feature's values; 2D heatmap for the FCs of MNIST digits.

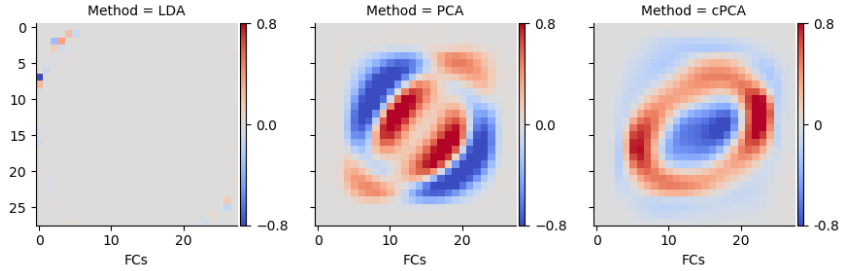
As general findings, LDA has more selected the features where either of the target cluster or the others has zero or little variance as the highly contributed features. This happened more frequently when the number of features is large. Also, PCA only has found the variations within the target cluster. On the other hand, ccPCA has found the features which make the target cluster unique but have enough variances.

C.1 MNIST Dataset

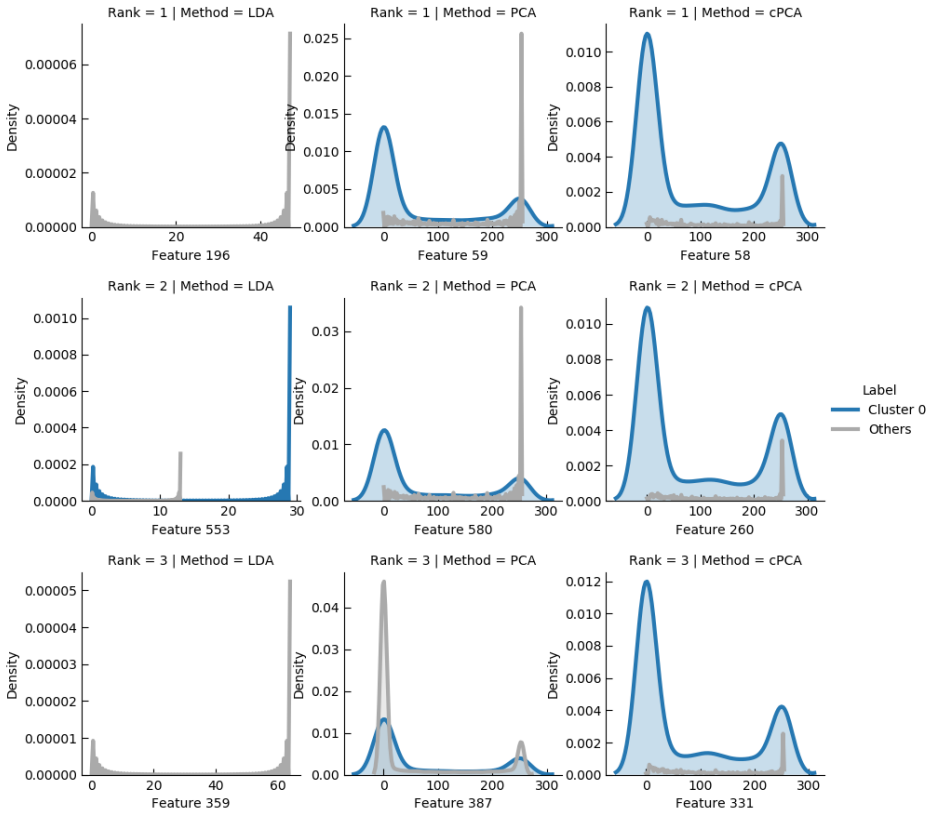
From Fig. C.1 to Fig. C.10, we show the results of the MNIST dataset. Generally, we can see that LDA selects the outside pixels as the features with high contributions because either of the target cluster or the others has no or little variance in the selected pixels. This is not preferable to understand each digit's characteristics. Also, PCA only shows the variations of strokes when drawing each digit. Therefore, we cannot understand why these digits are unique when compared with others. On the other hand, ccPCA results clearly show the strokes contrasting the target digit from the others (e.g., the pixels on the top right in Digit 5).



(a) 1D DR result

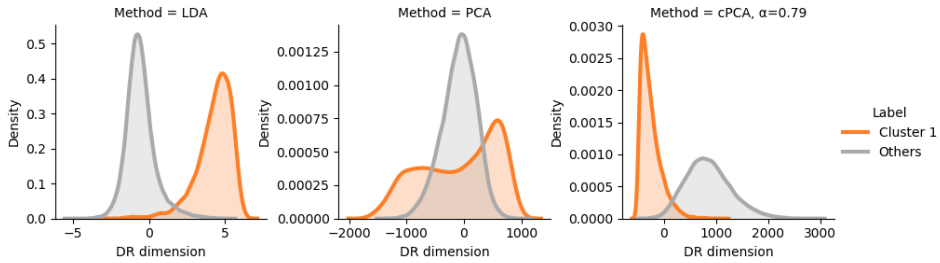


(b) Feature contributions

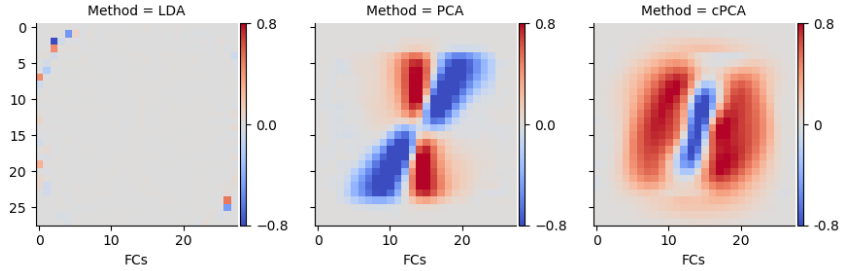


(c) Distributions of the top 3 features' values

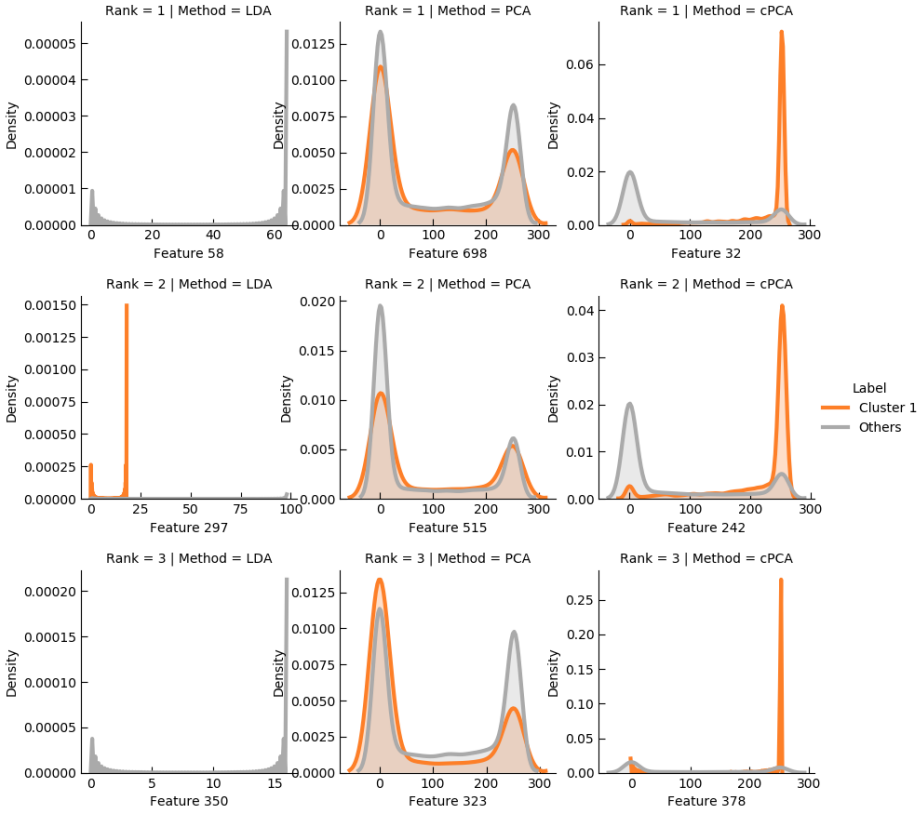
Figure C.1: MNIST Digit 0



(a) 1D DR result

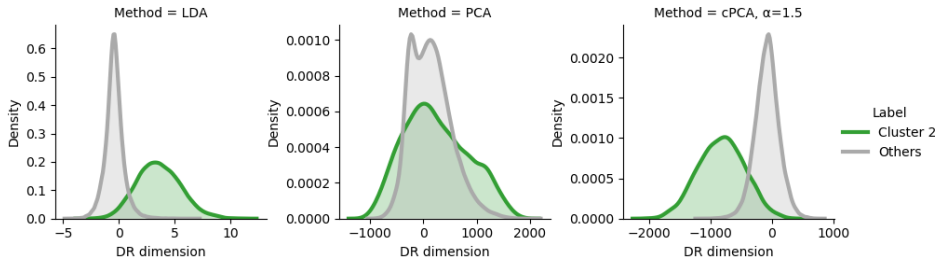


(b) Feature contributions

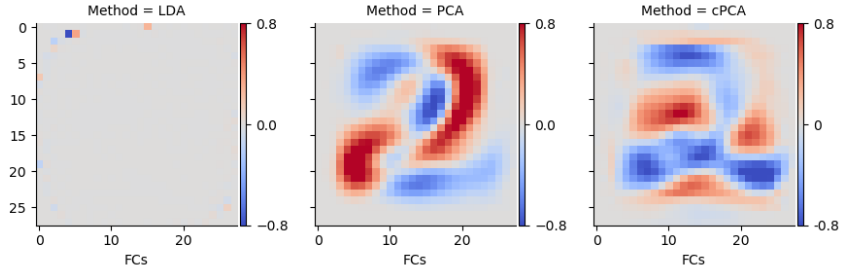


(c) Distributions of the top 3 features' values

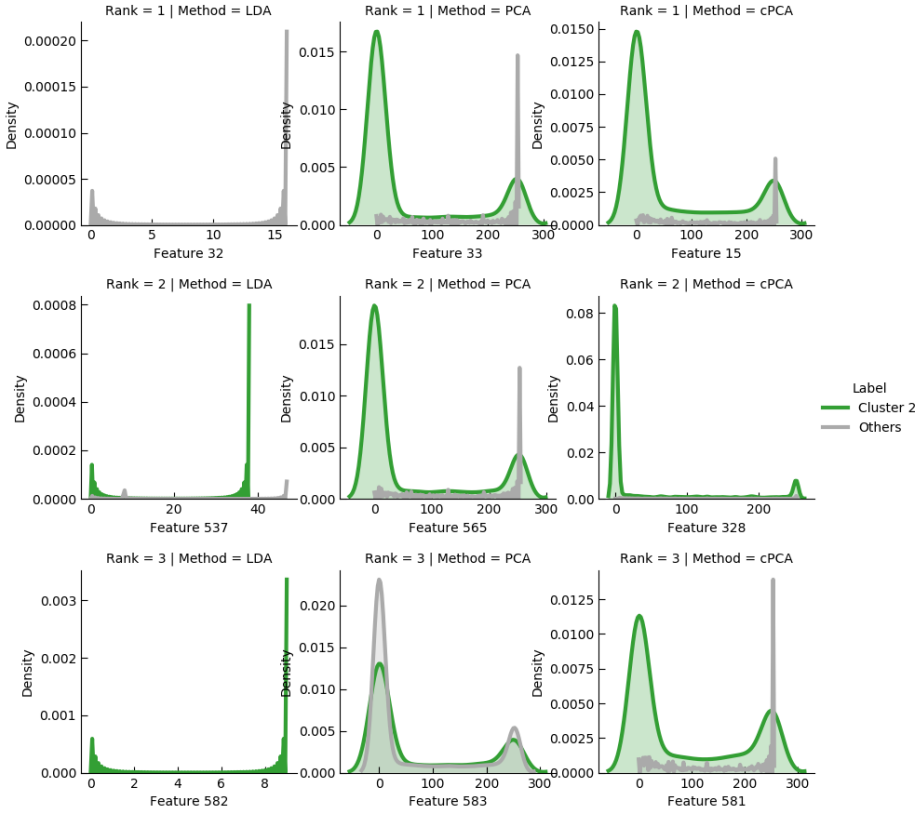
Figure C.2: MNIST Digit 1



(a) 1D DR result

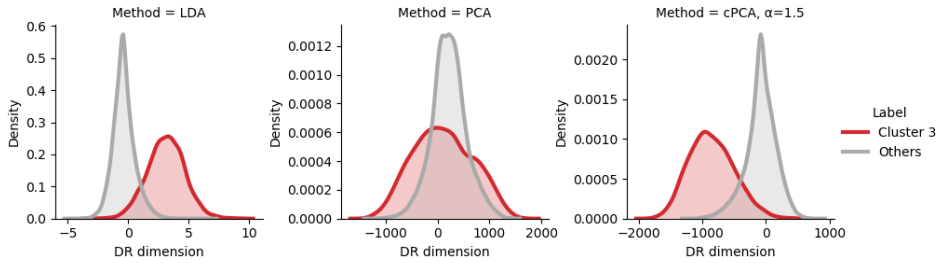


(b) Feature contributions

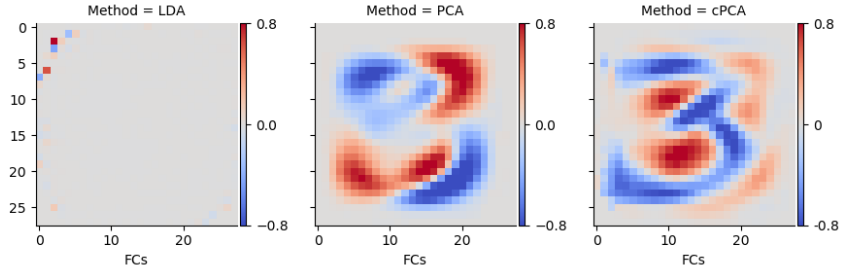


(c) Distributions of the top 3 features' values

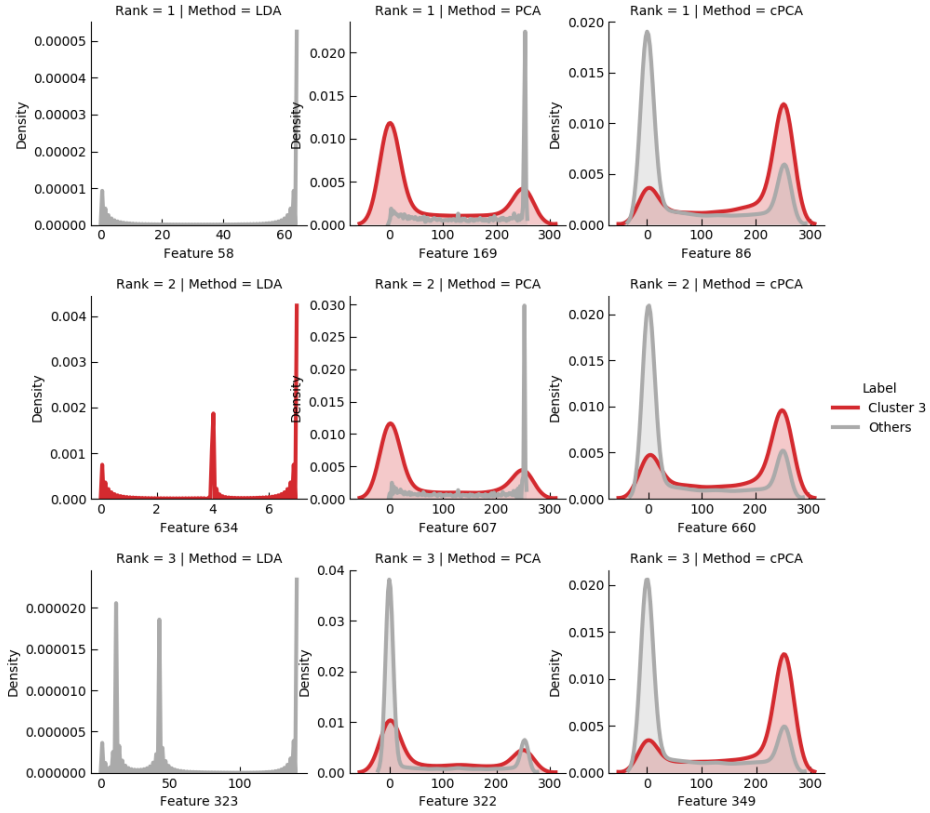
Figure C.3: MNIST Digit 2



(a) 1D DR result

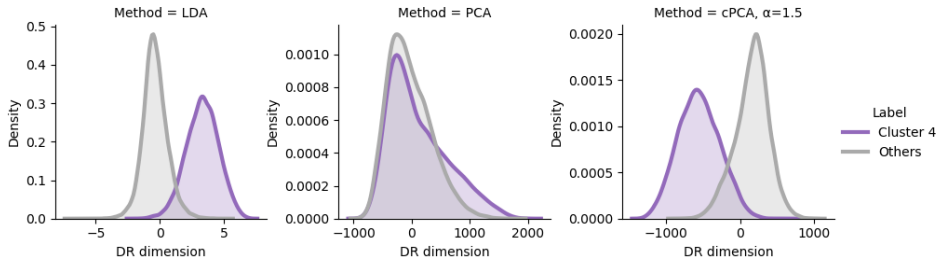


(b) Feature contributions

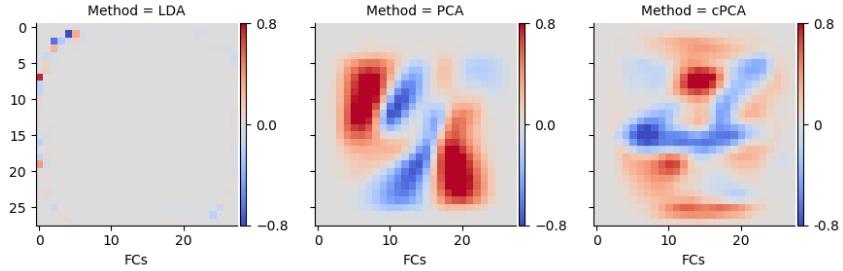


(c) Distributions of the top 3 features' values

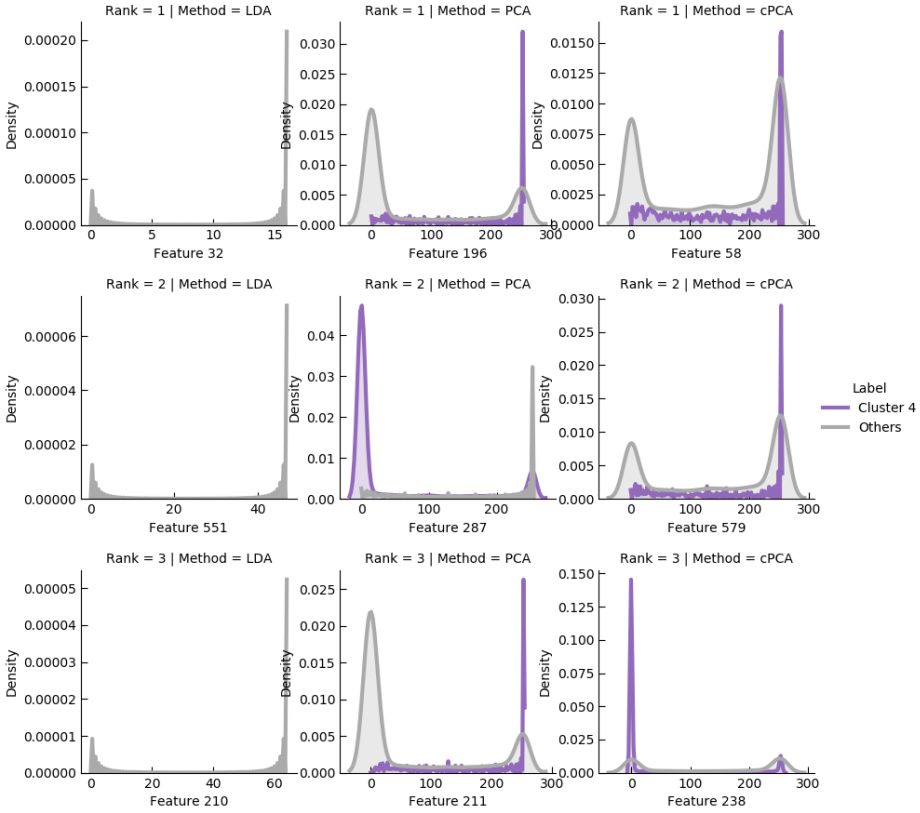
Figure C.4: MNIST Digit 3



(a) 1D DR result

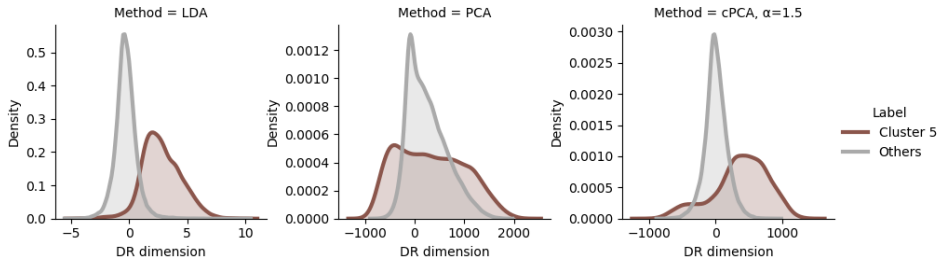


(b) Feature contributions

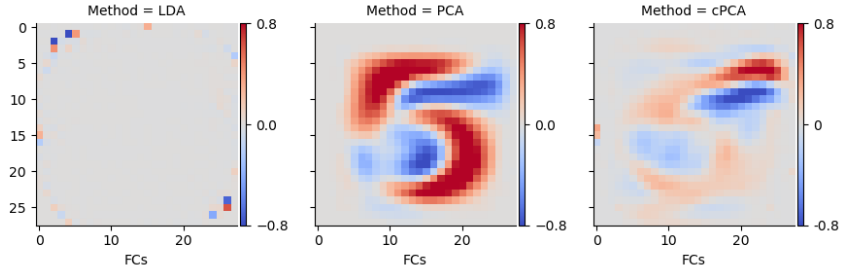


(c) Distributions of the top 3 features' values

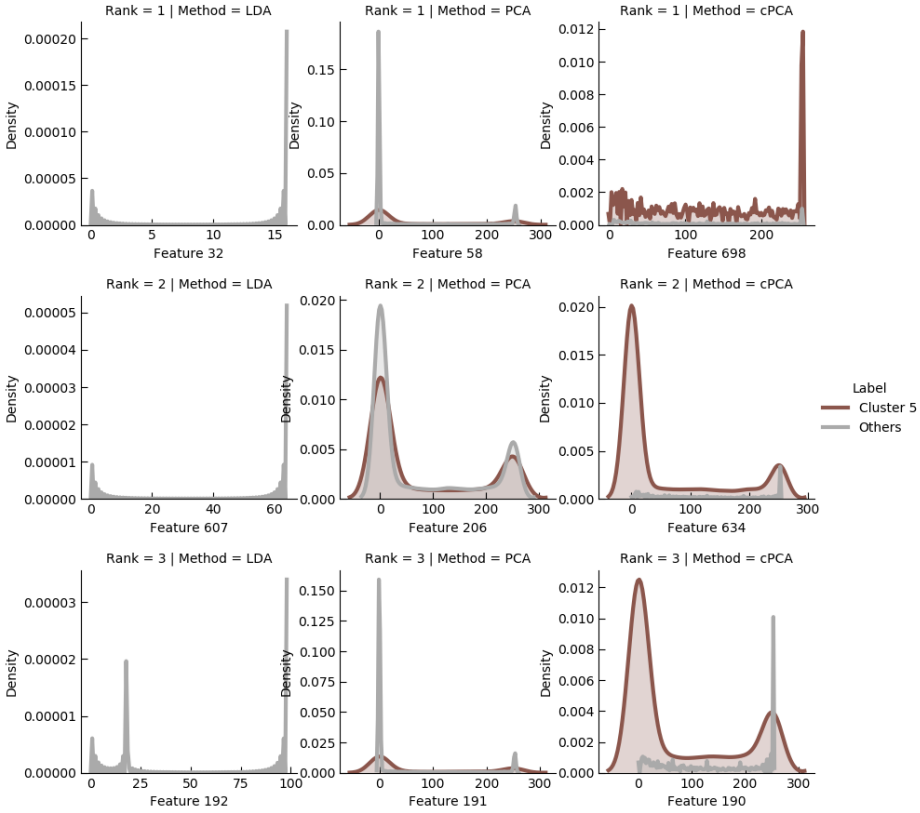
Figure C.5: MNIST Digit 4



(a) 1D DR result

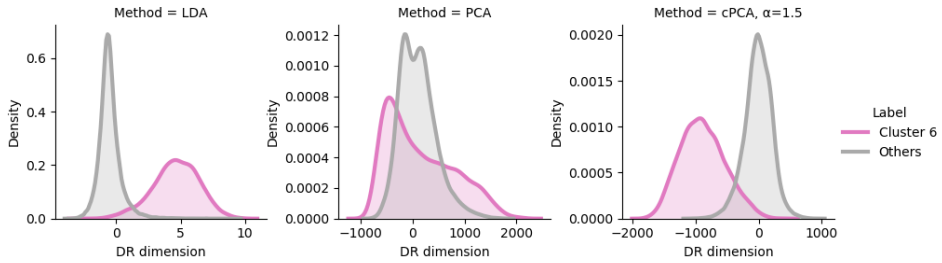


(b) Feature contributions

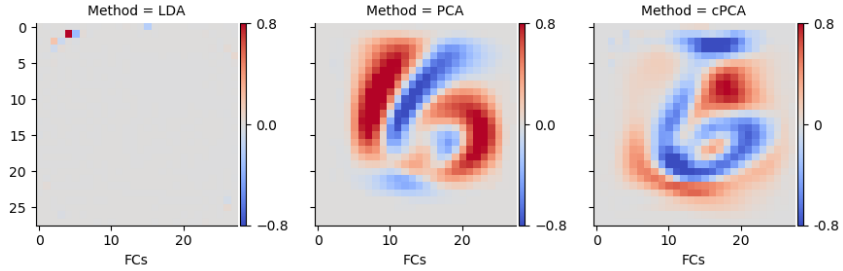


(c) Distributions of the top 3 features' values

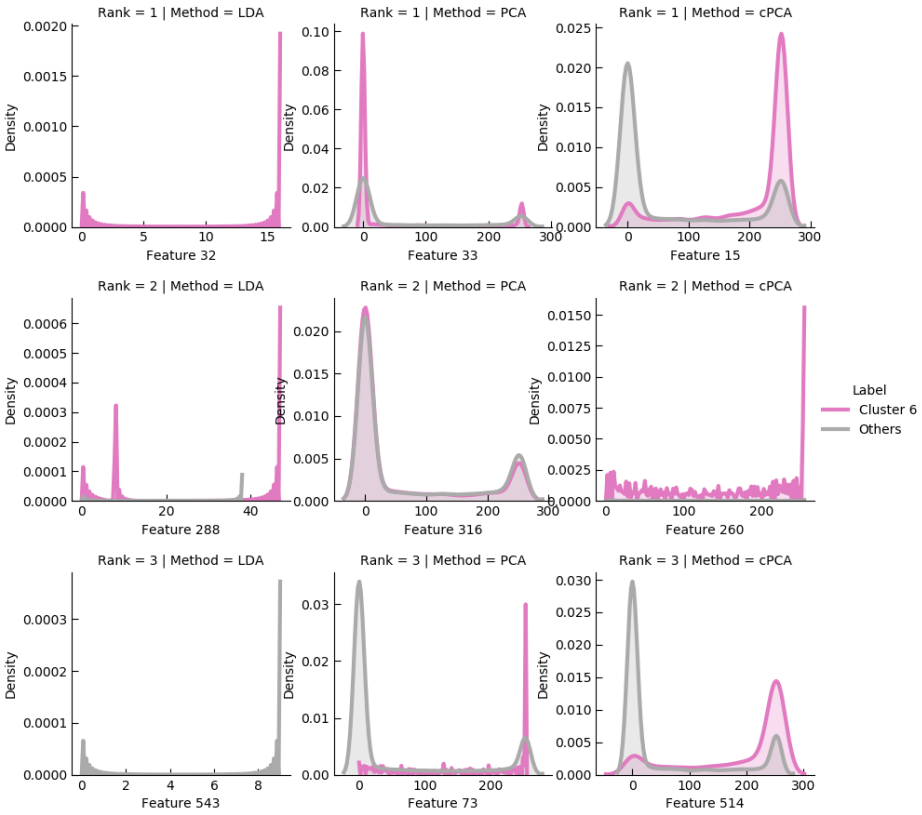
Figure C.6: MNIST Digit 5



(a) 1D DR result

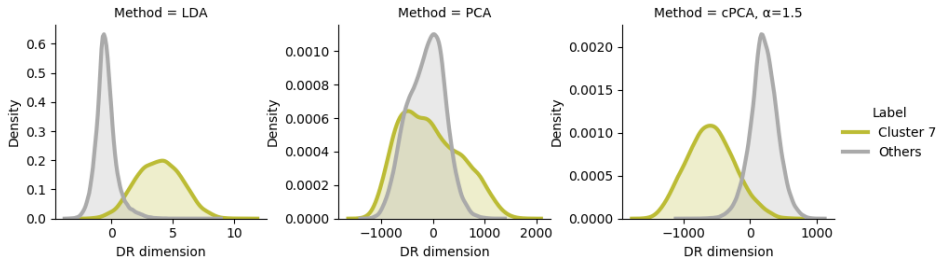


(b) Feature contributions

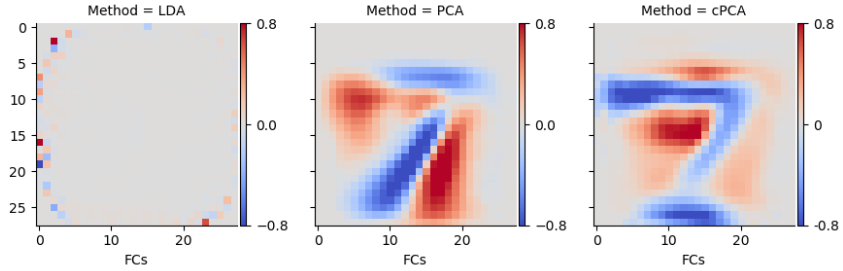


(c) Distributions of the top 3 features' values

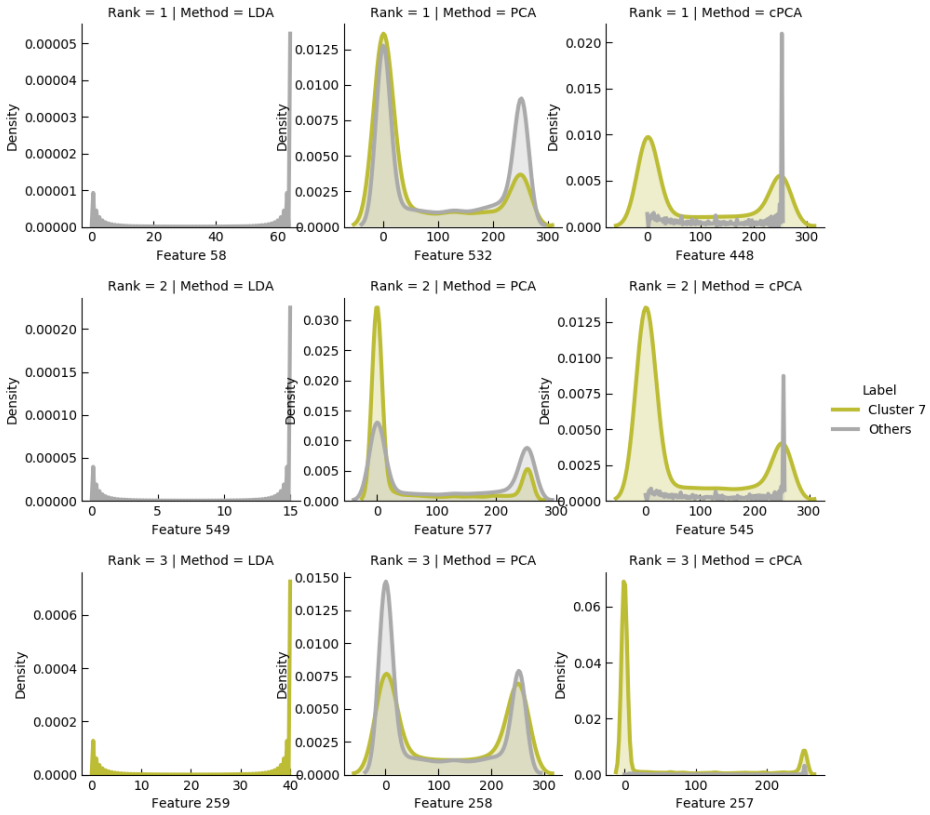
Figure C.7: MNIST Digit 6



(a) 1D DR result

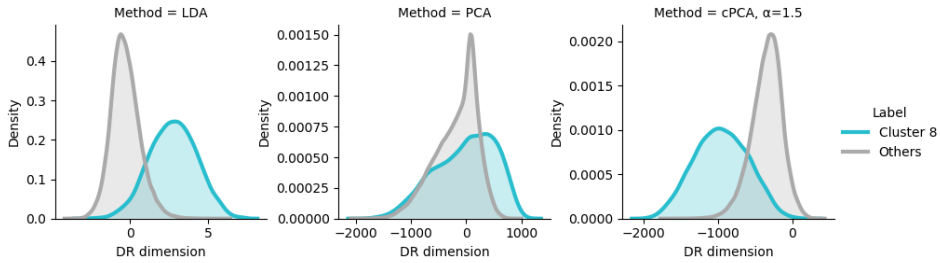


(b) Feature contributions

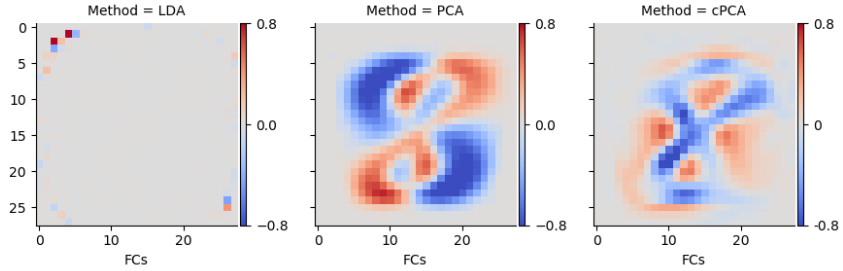


(c) Distributions of the top 3 features' values

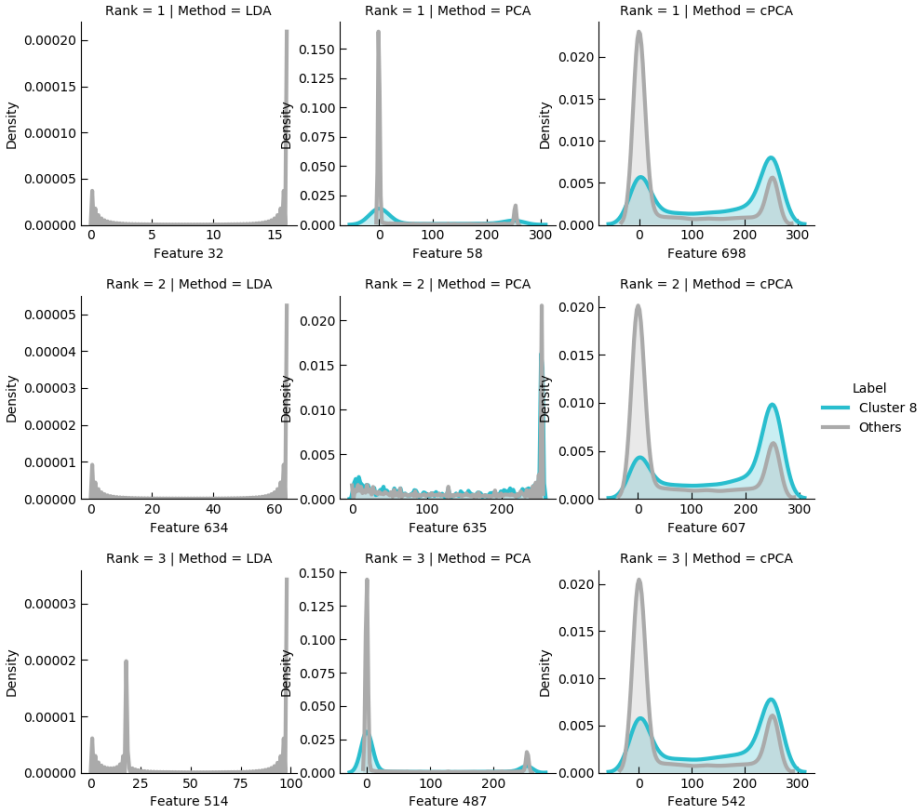
Figure C.8: MNIST Digit 7



(a) 1D DR result

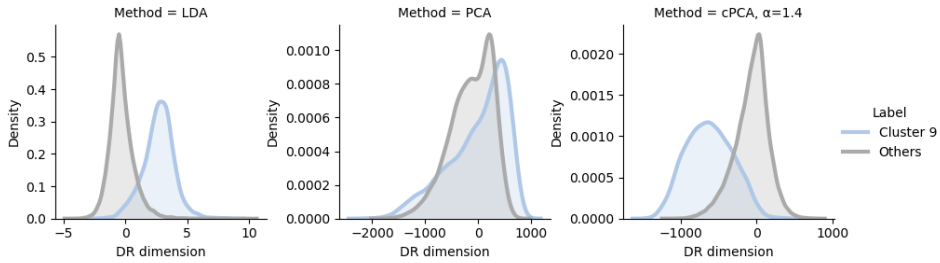


(b) Feature contributions

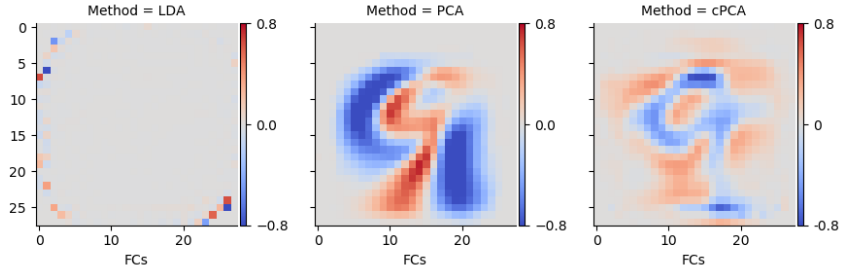


(c) Distributions of the top 3 features' values

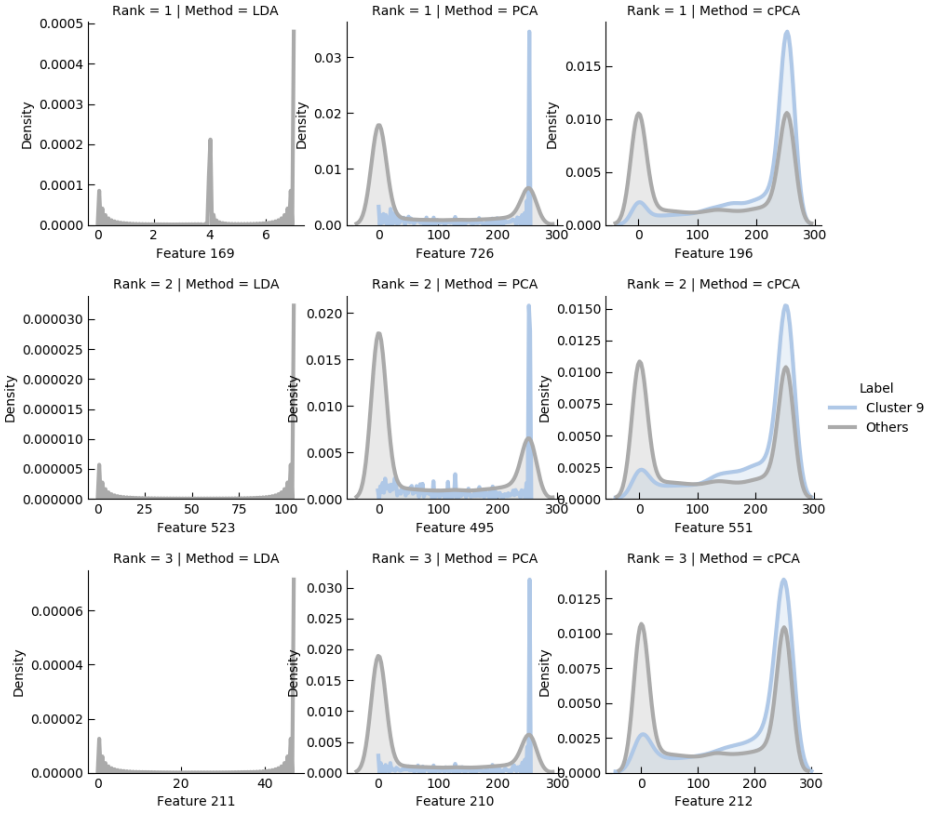
Figure C.9: MNIST Digit 8



(a) 1D DR result



(b) Feature contributions

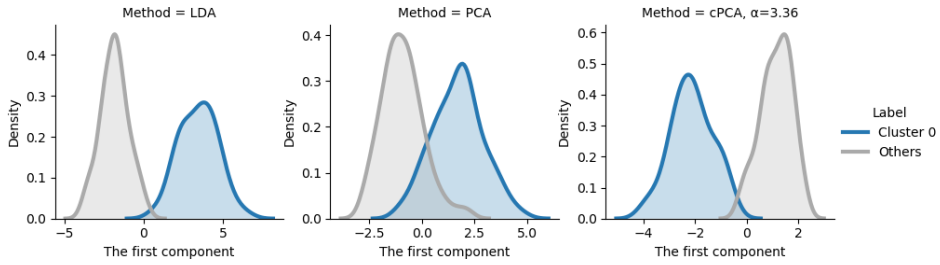


(c) Distributions of the top 3 features' values

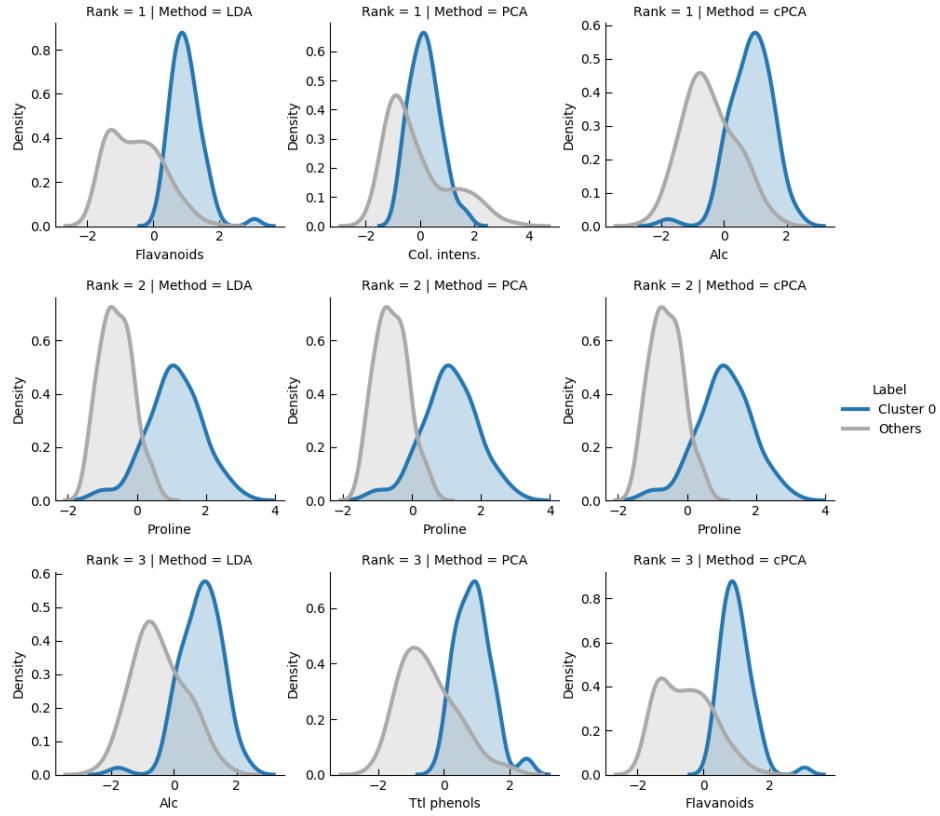
Figure C.10: MNIST Digit 9

C.2 Wine Recognition Dataset

From Fig. C.11 to Fig. C.13, we show the results of the Wine Recognition dataset. Because this dataset has only 13 features and each cluster has similar variances in each feature, LDA and ccPCA tend to have similar results.

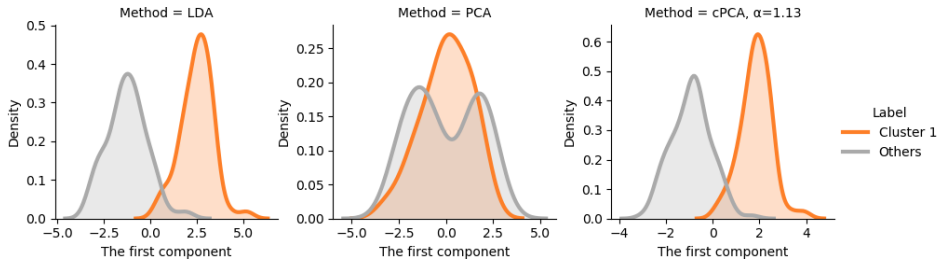


(a) 1D DR result

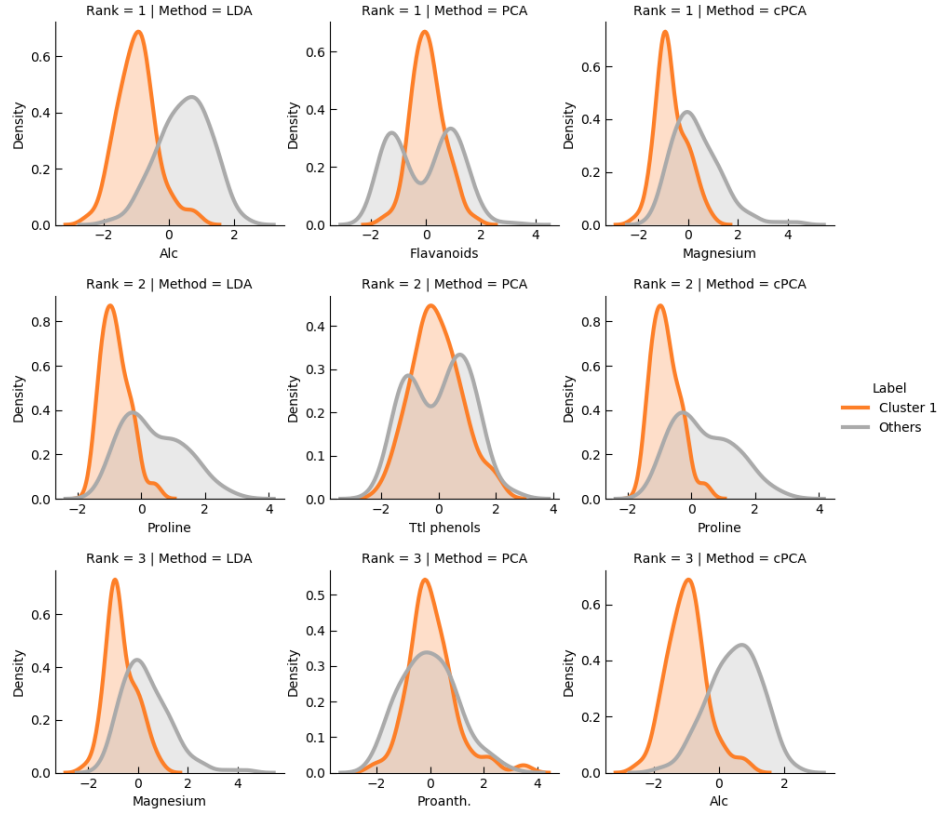


(b) Distributions of the top 3 features' values

Figure C.11: Wine Recognition: Cluster 0

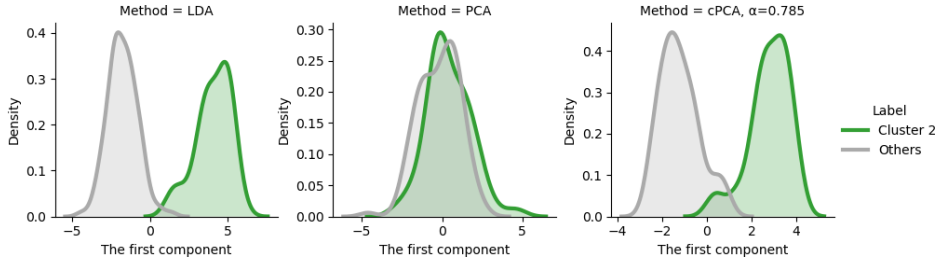


(a) 1D DR result

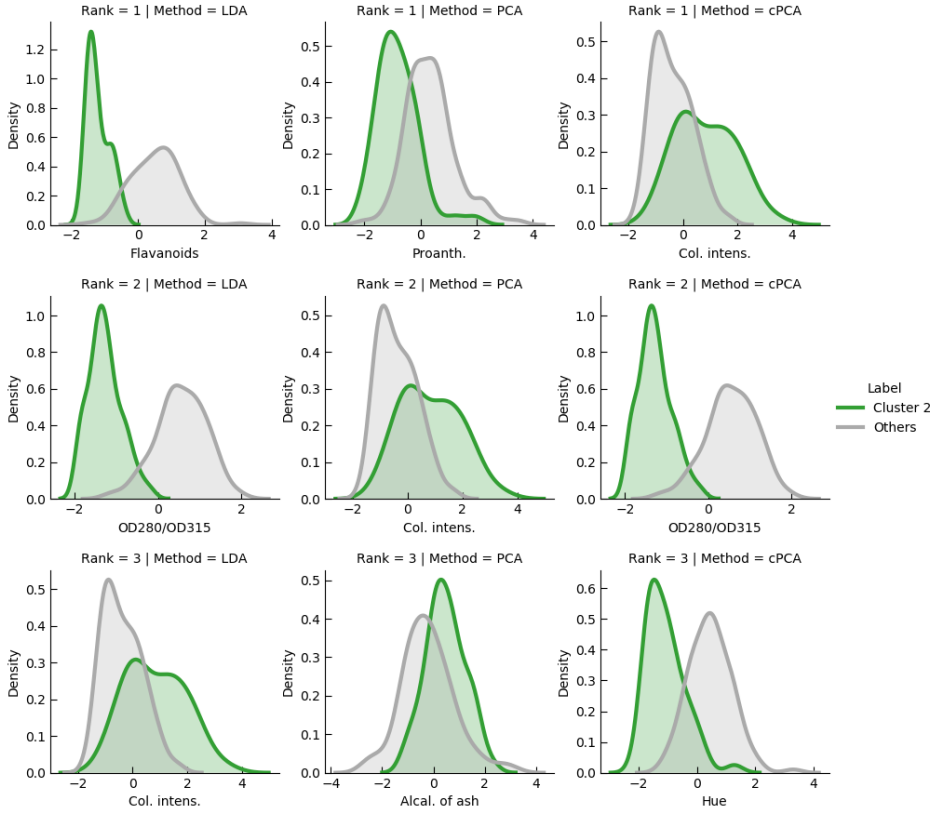


(b) Distributions of the top 3 features' values

Figure C.12: Wine Recognition: Cluster 1



(a) 1D DR result

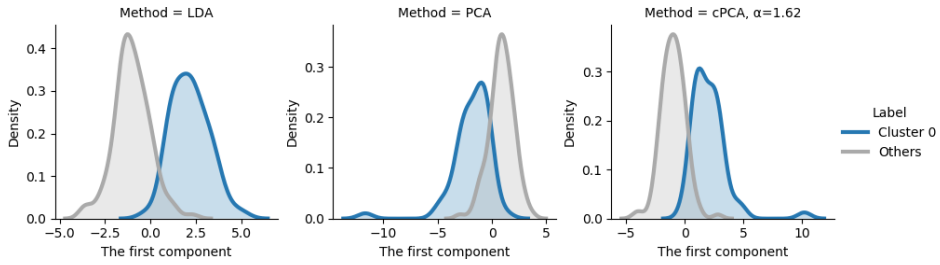


(b) Distributions of the top 3 features' values

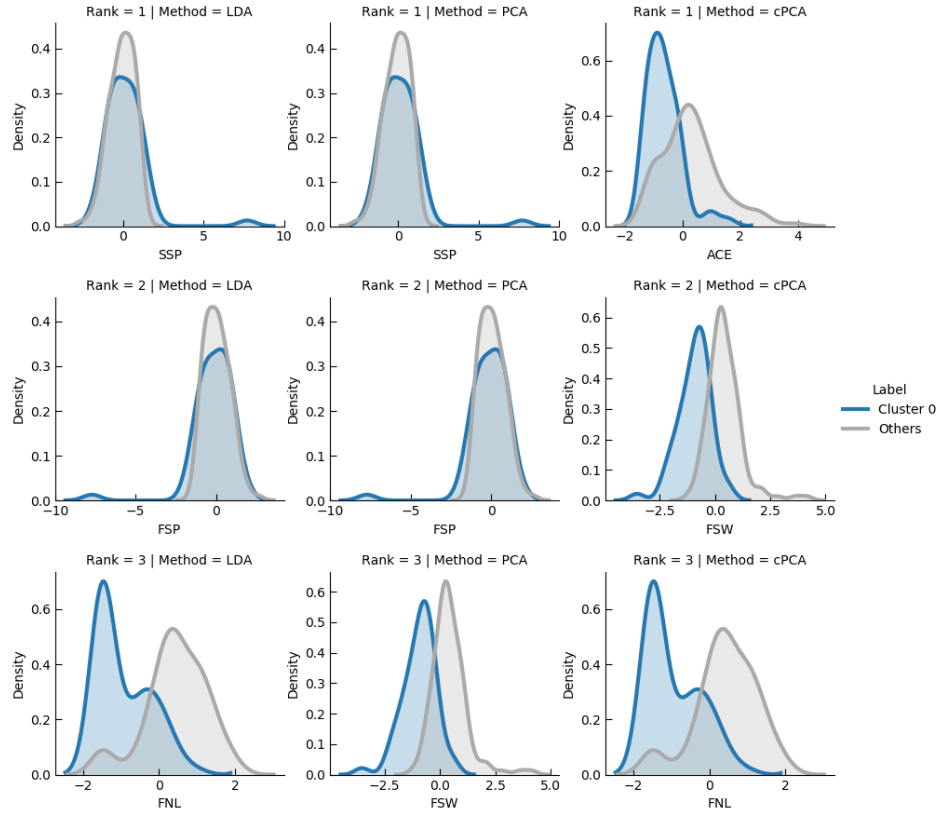
Figure C.13: Wine Recognition: Cluster 2

C.3 Tennis Major Tournament Match Statistics Dataset

From Fig. C.14 to Fig. C.18, we show the results of the Tennis Major Tournament Match Statistics dataset. We can see that ccPCA tends to select the features which have better separation when compared with the features selected by LDA or PCA.

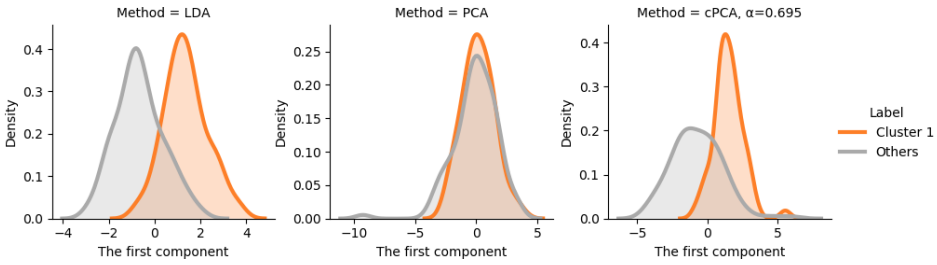


(a) 1D DR result

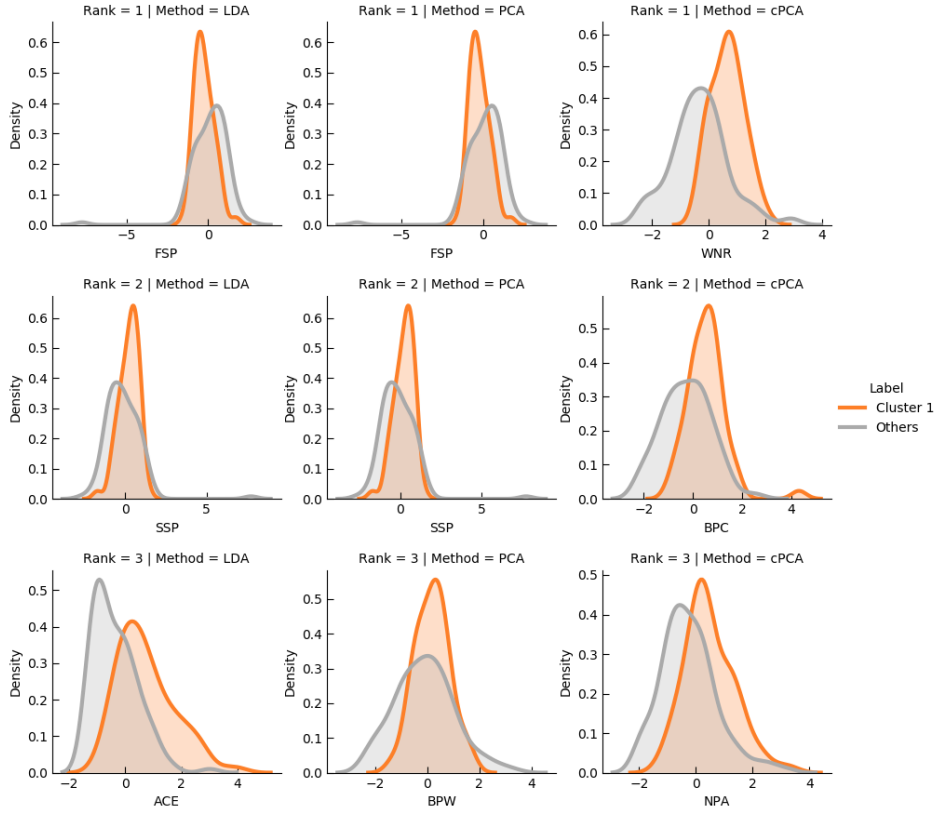


(b) Distributions of the top 3 features' values

Figure C.14: Tennis Major Tournament Match Statistics: Cluster 0

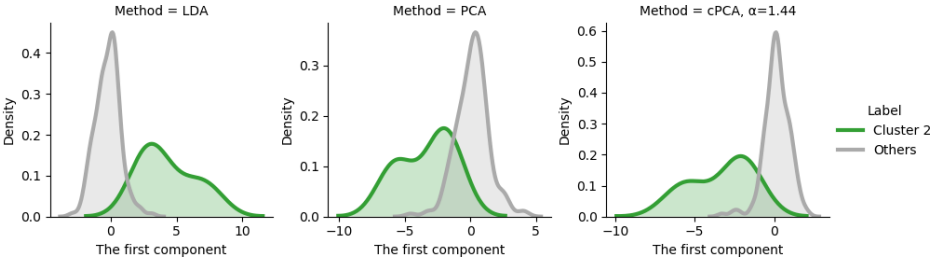


(a) 1D DR result

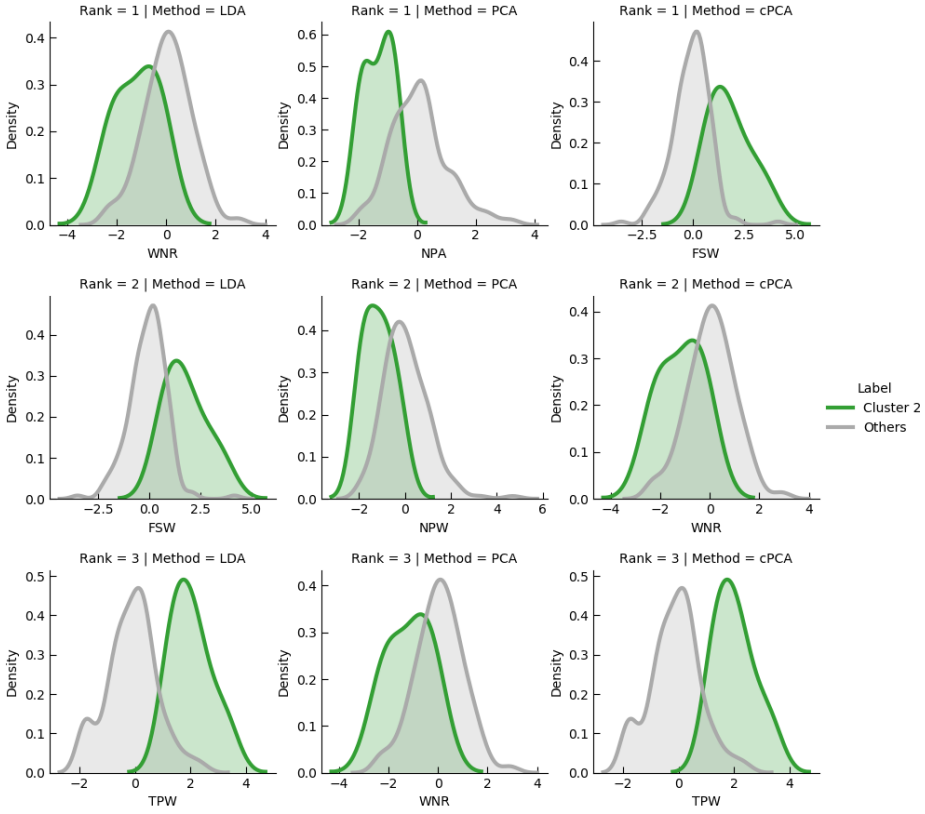


(b) Distributions of the top 3 features' values

Figure C.15: Tennis Major Tournament Match Statistics: Cluster 1

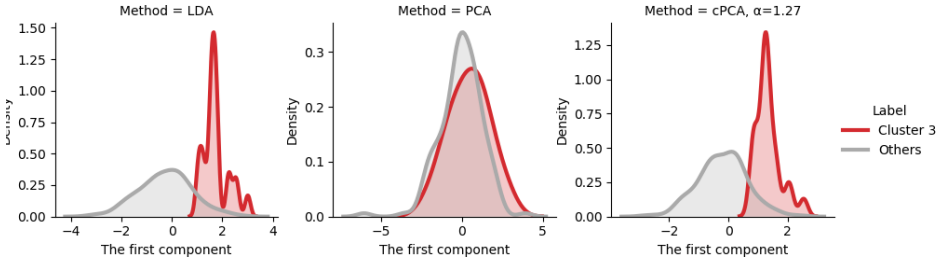


(a) 1D DR result

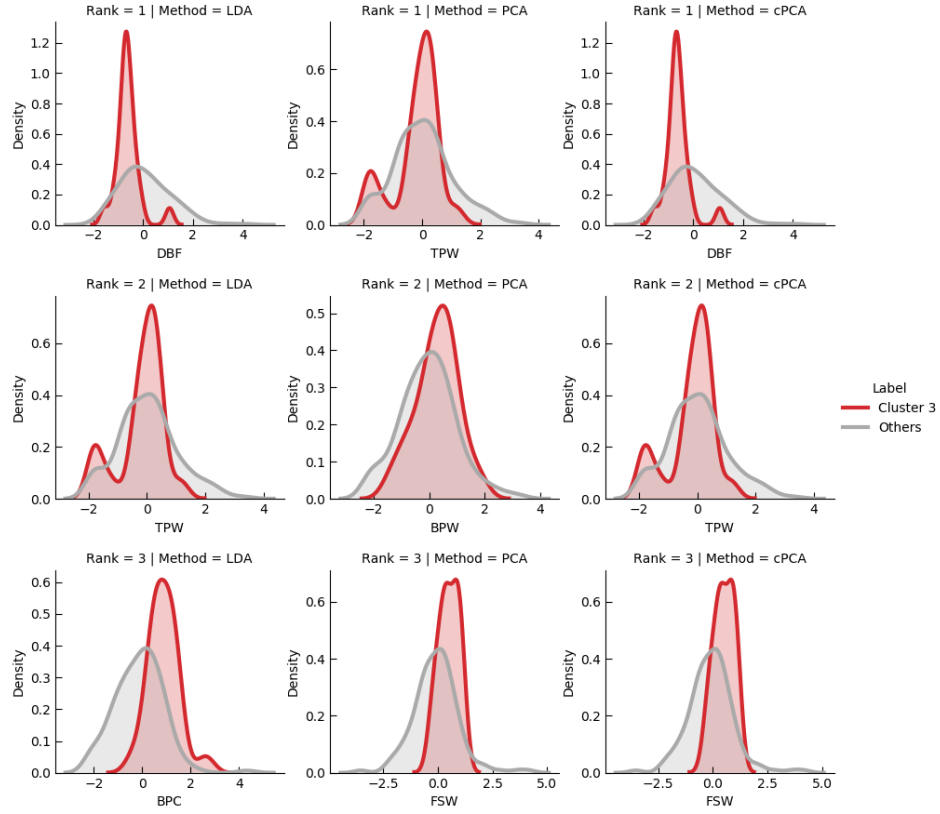


(b) Distributions of the top 3 features' values

Figure C.16: Tennis Major Tournament Match Statistics: Cluster 2

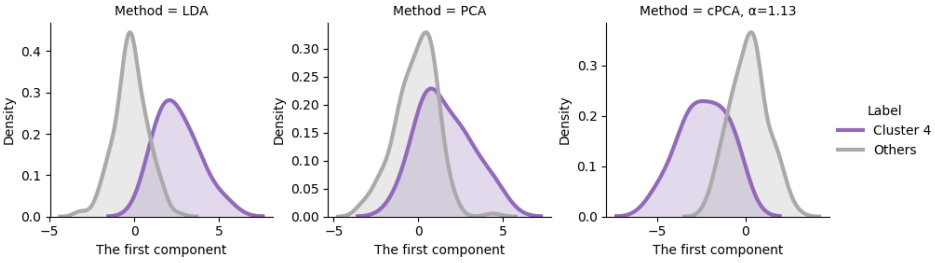


(a) 1D DR result

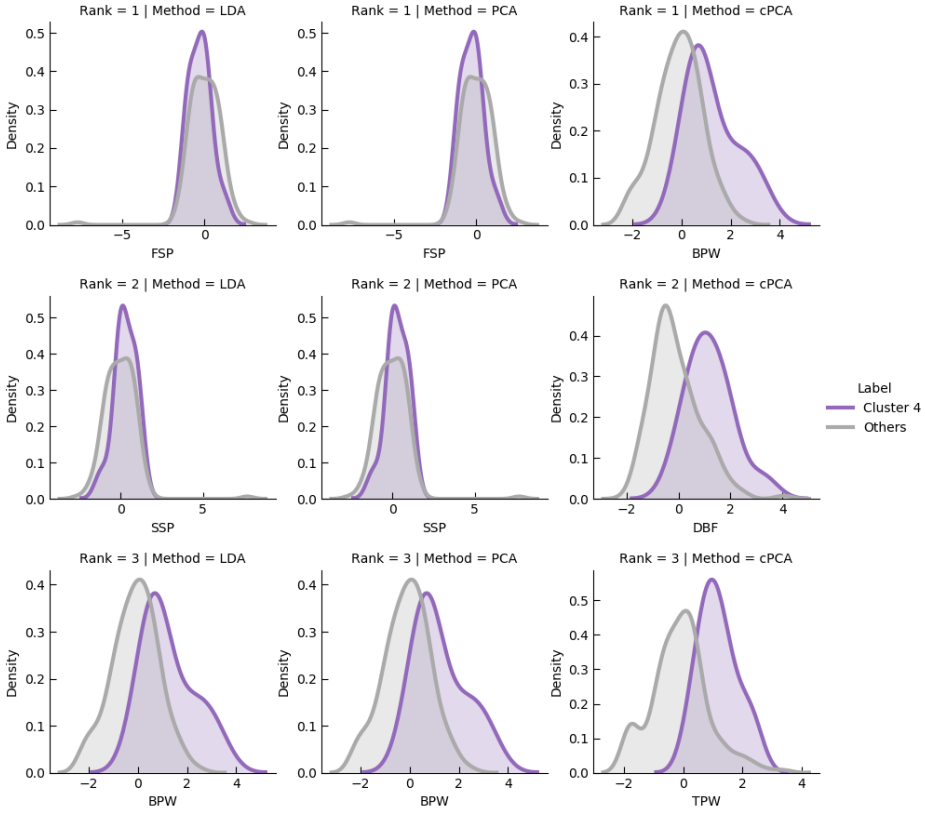


(b) Distributions of the top 3 features' values

Figure C.17: Tennis Major Tournament Match Statistics: Cluster 3



(a) 1D DR result

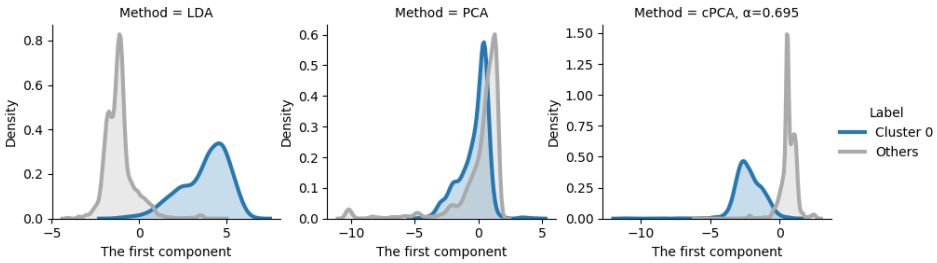


(b) Distributions of the top 3 features' values

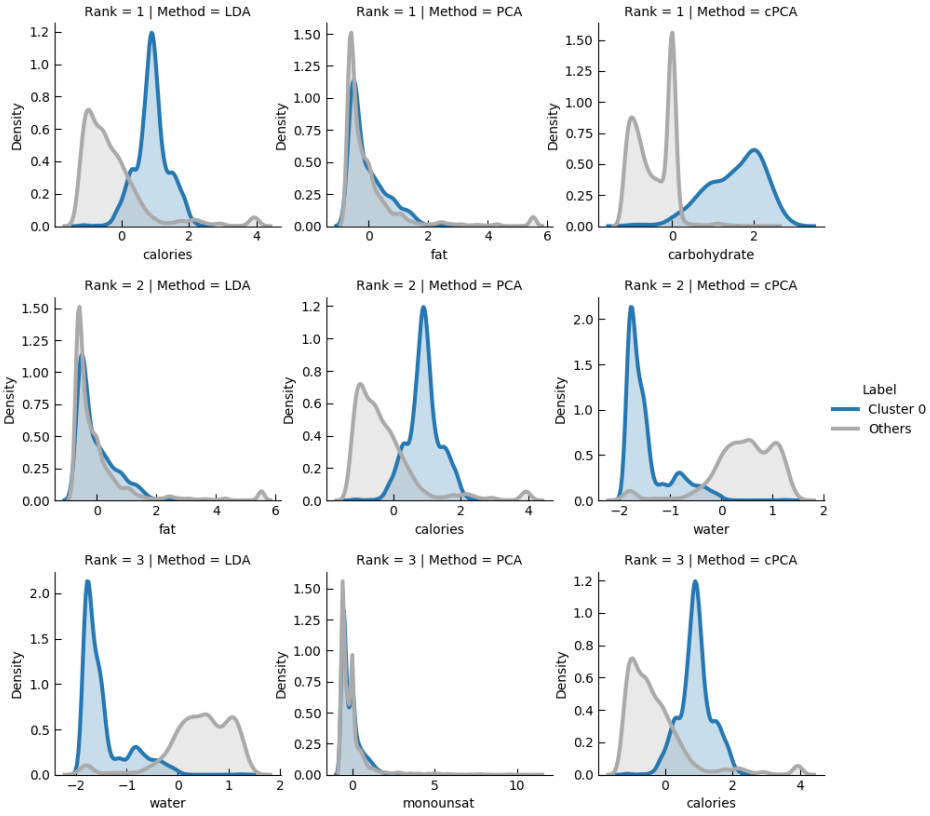
Figure C.18: Tennis Major Tournament Match Statistics: Cluster 4

C.4 Nutrients Dataset

From Fig. C.19 to Fig. C.27, we show the results of the Nutrients dataset. As described in Sect. B.4, because a few features are dominant to distinguish or contrast each target cluster, all LDA, PCA, and ccPCA have similar results in some of the clusters.

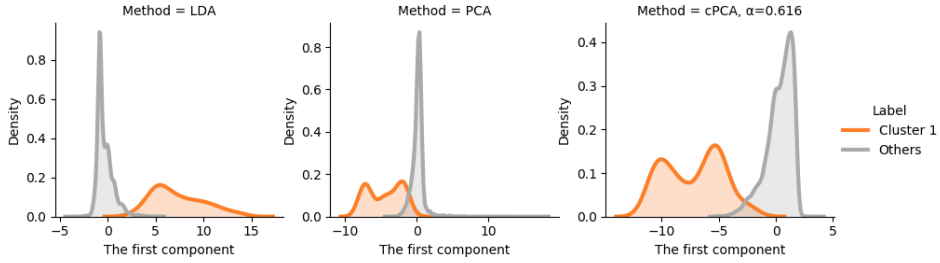


(a) 1D DR result

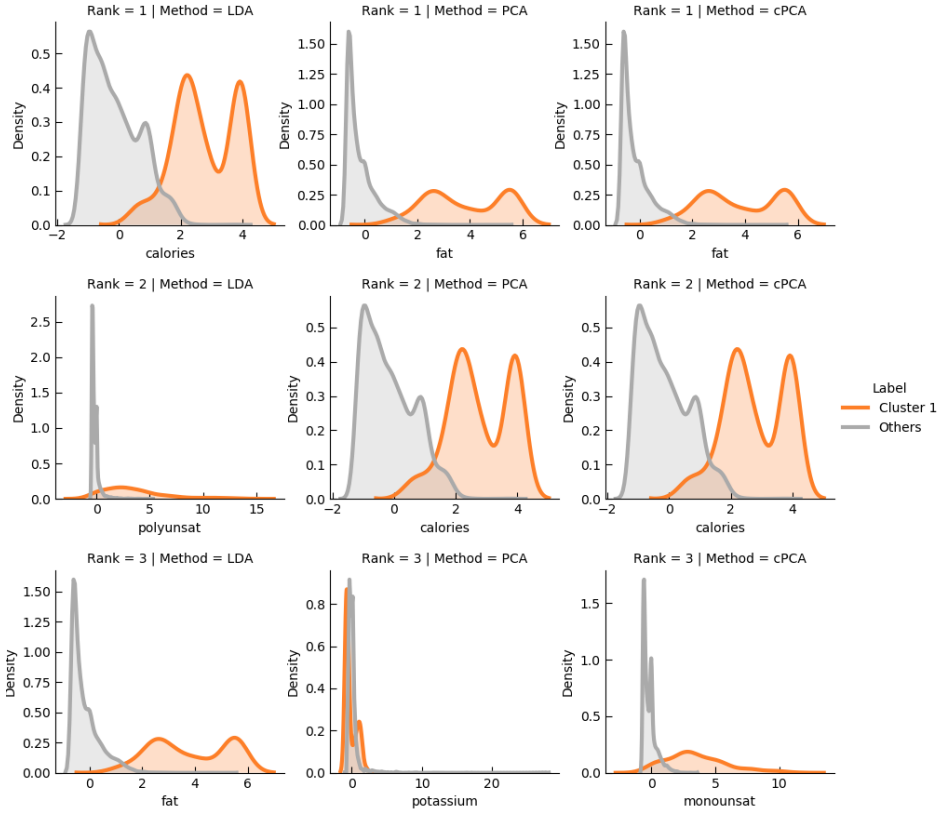


(b) Distributions of the top 3 features' values

Figure C.19: Nutrients: Cluster 0

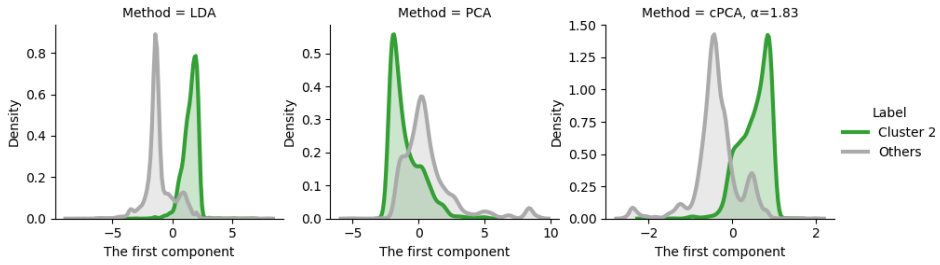


(a) 1D DR result

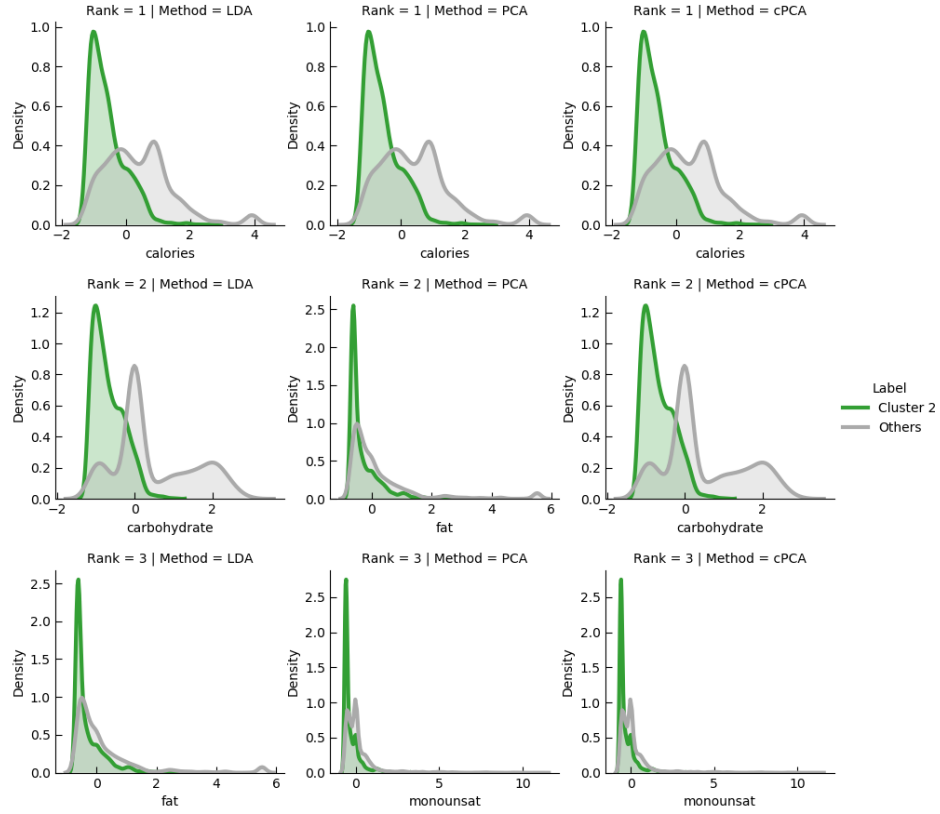


(b) Distributions of the top 3 features' values

Figure C.20: Nutrients: Cluster 1

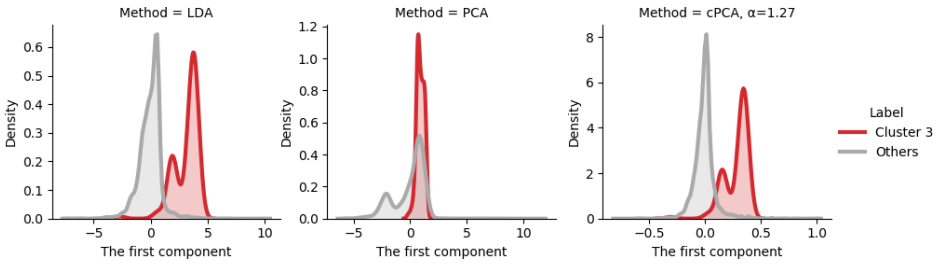


(a) 1D DR result

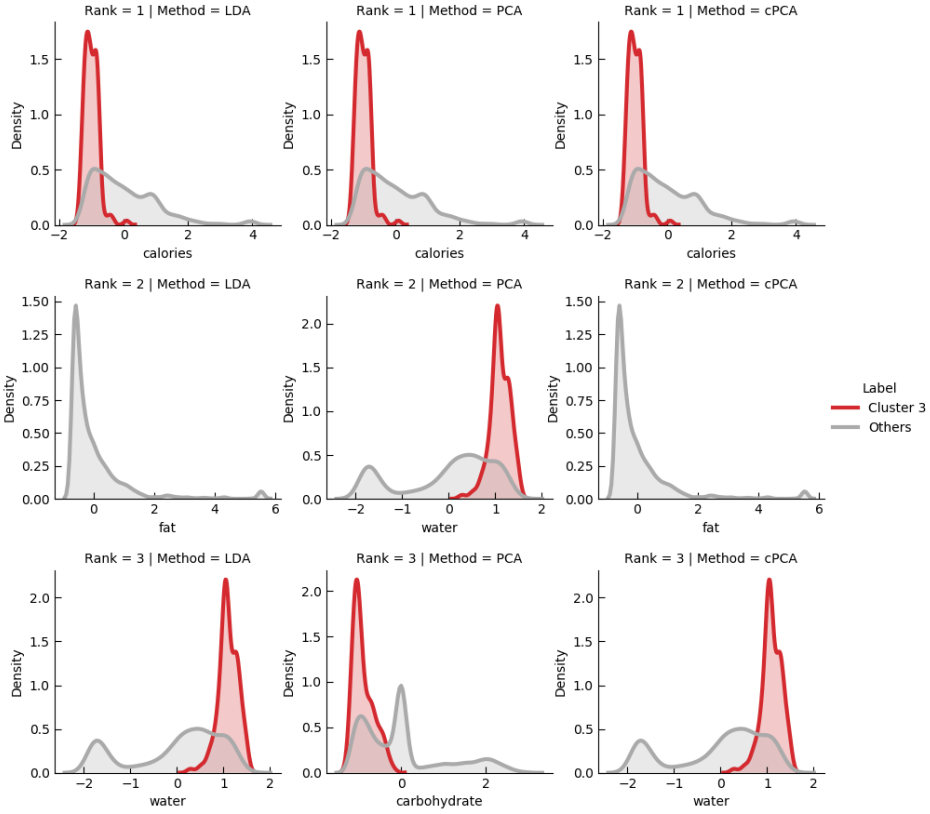


(b) Distributions of the top 3 features' values

Figure C.21: Nutrients: Cluster 2

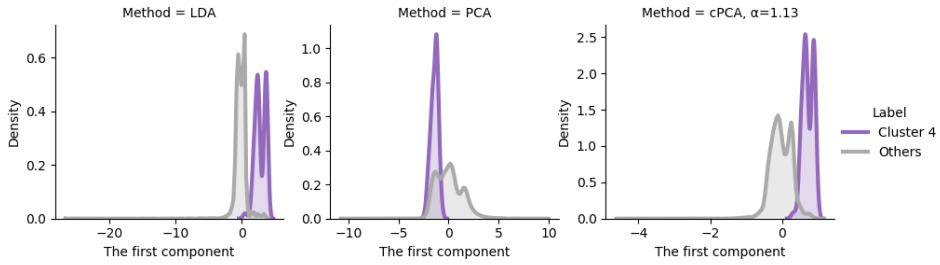


(a) 1D DR result

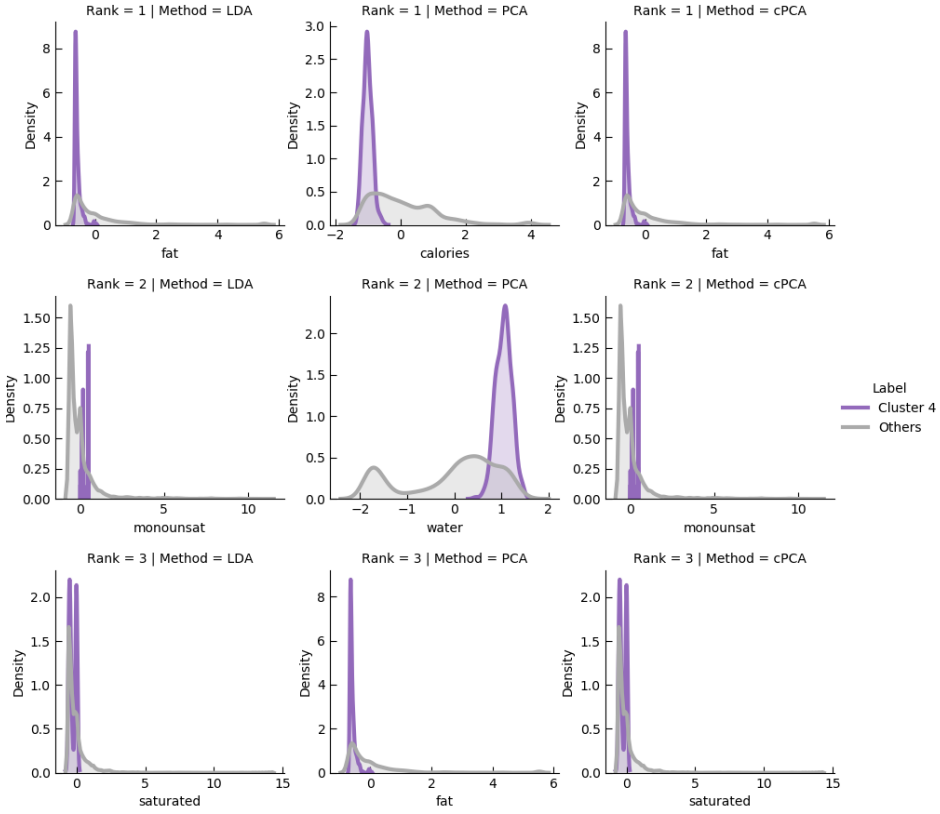


(b) Distributions of the top 3 features' values

Figure C.22: Nutrients: Cluster 3

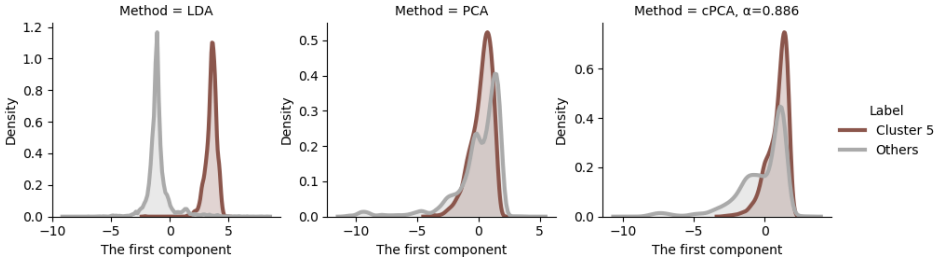


(a) 1D DR result

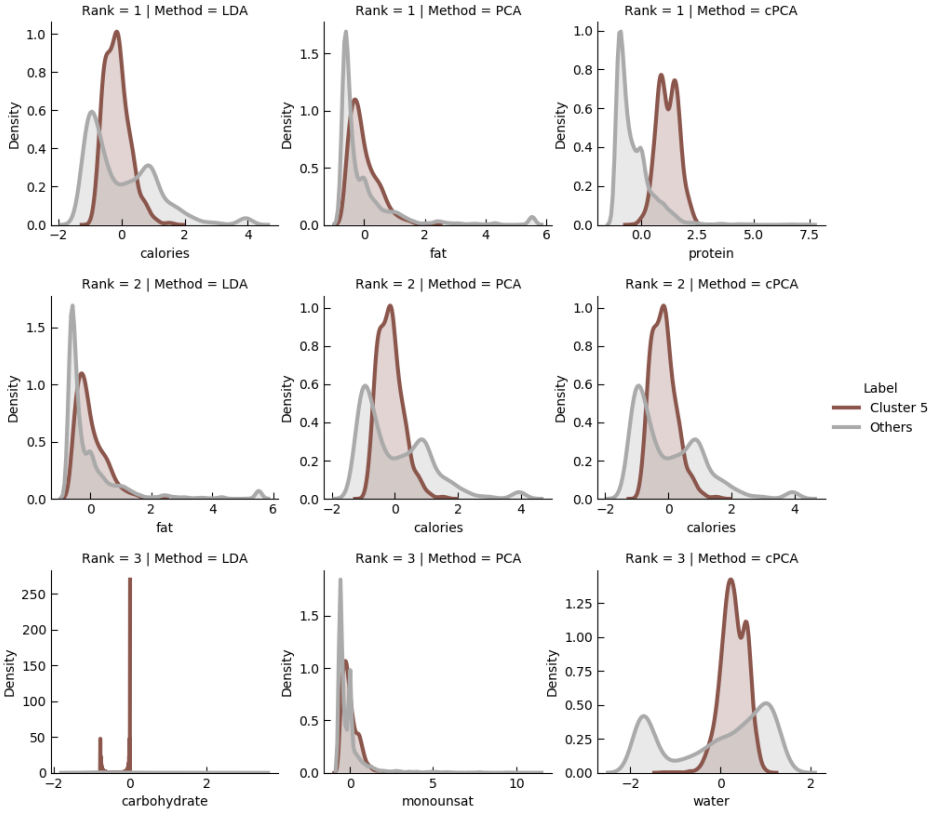


(b) Distributions of the top 3 features' values

Figure C.23: Nutrients: Cluster 4

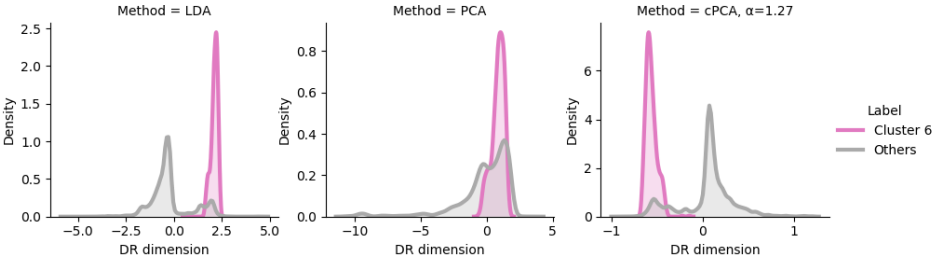


(a) 1D DR result

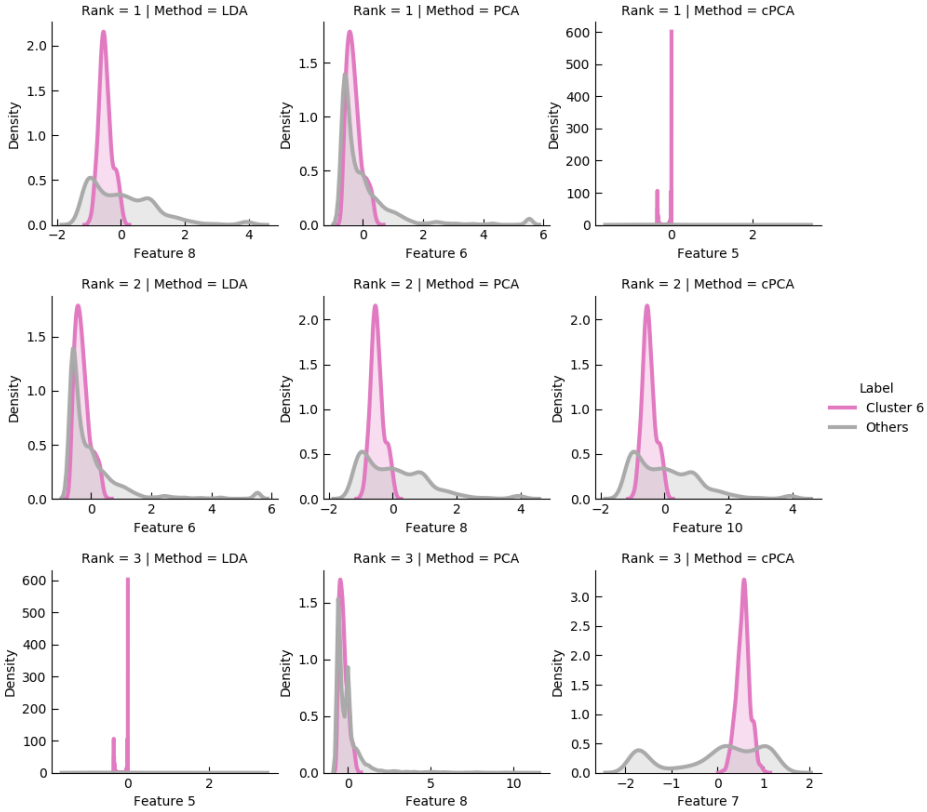


(b) Distributions of the top 3 features' values

Figure C.24: Nutrients: Cluster 5

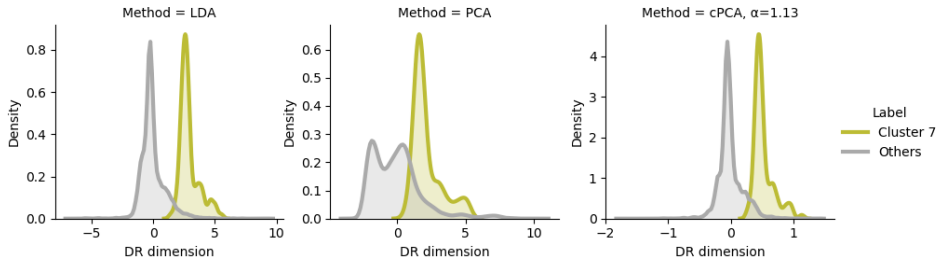


(a) 1D DR result

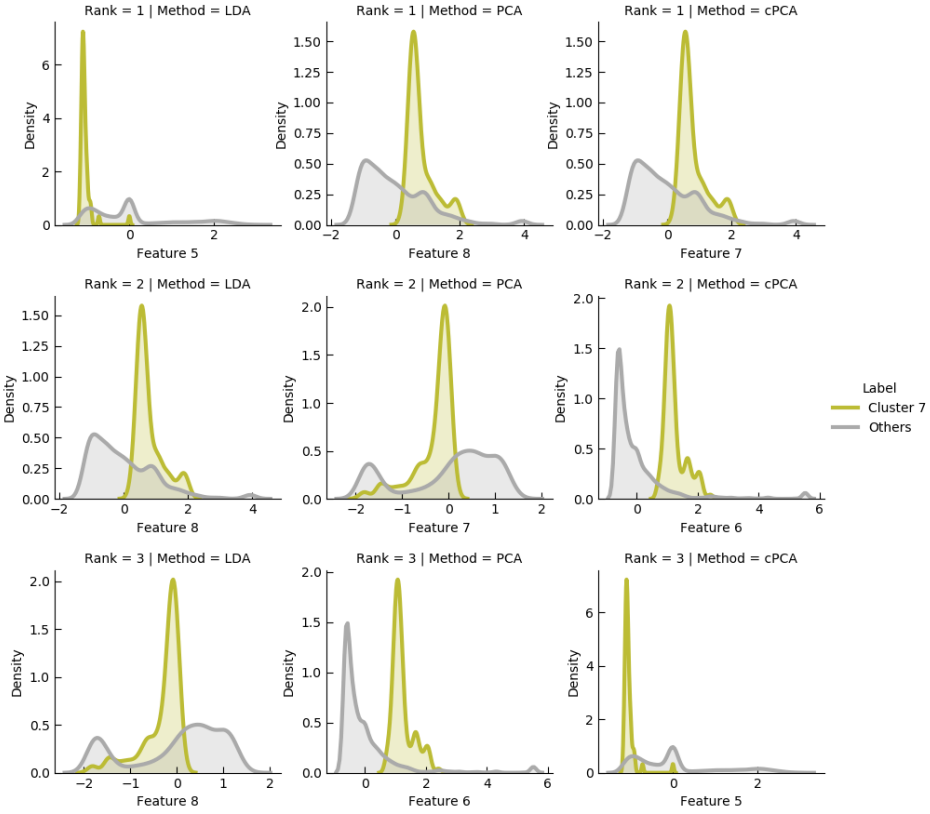


(b) Distributions of the top 3 features' values

Figure C.25: Nutrients: Cluster 6

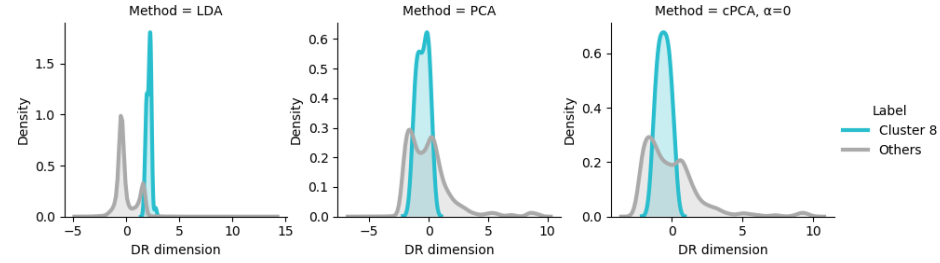


(a) 1D DR result

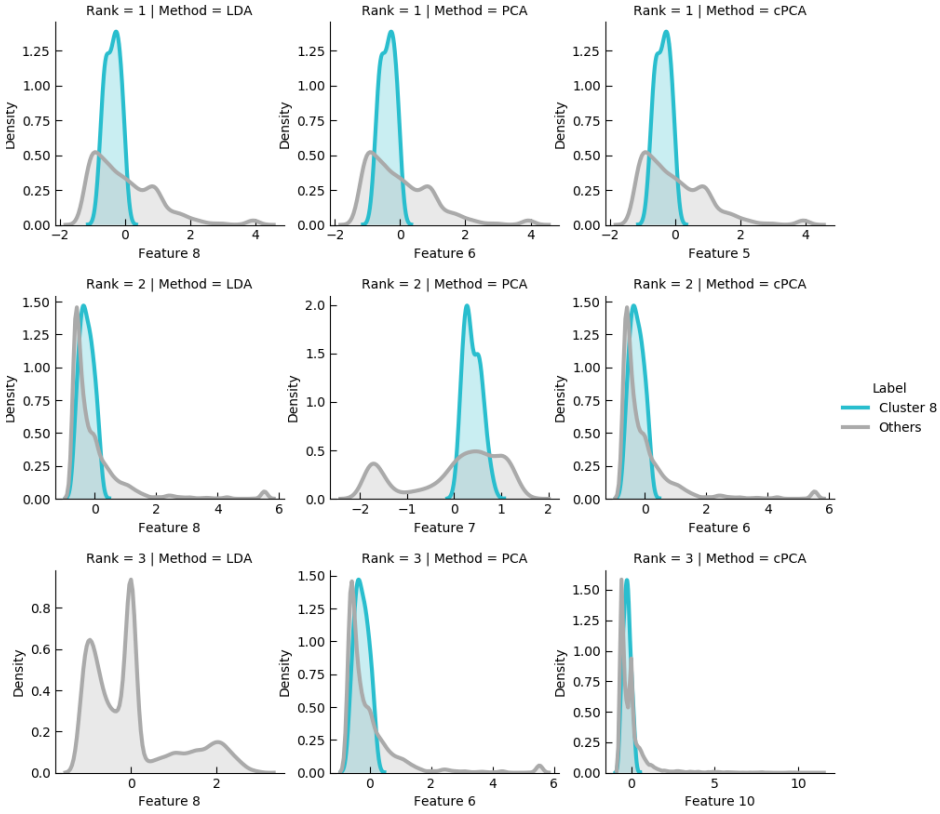


(b) Distributions of the top 3 features' values

Figure C.26: Nutrients: Cluster 7



(a) 1D DR result

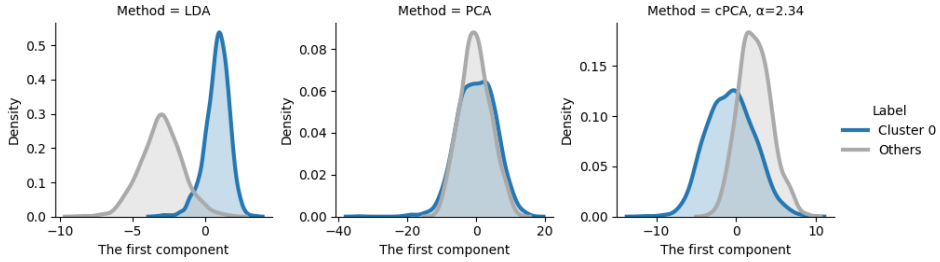


(b) Distributions of the top 3 features' values

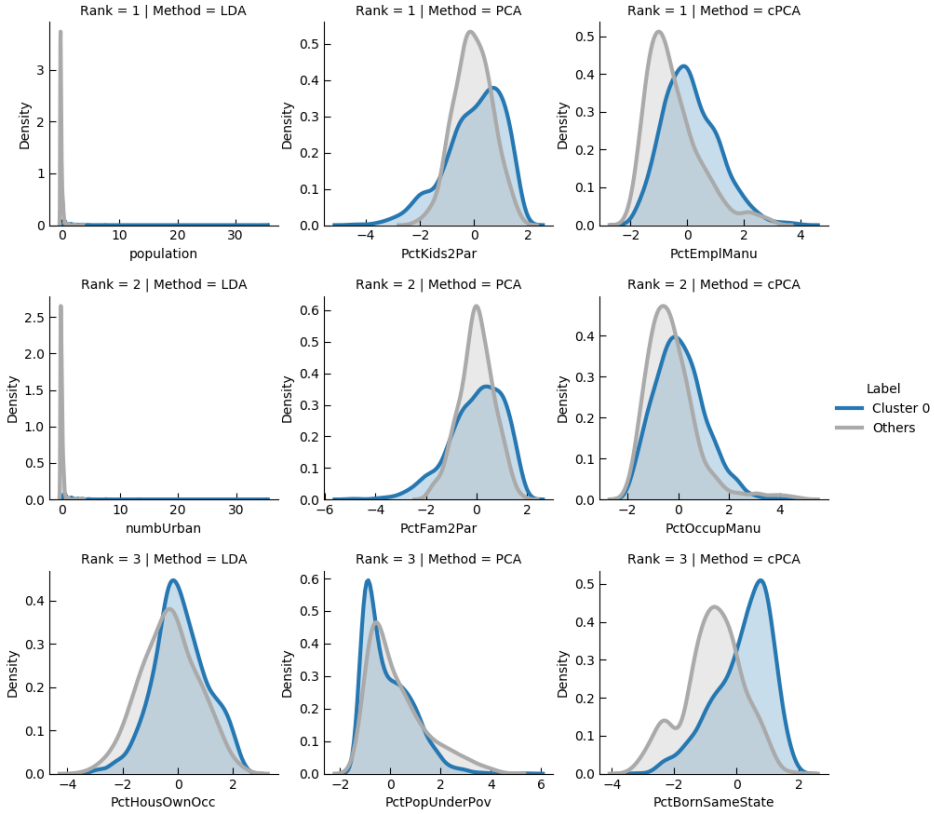
Figure C.27: Nutrients: Cluster 8

C.5 Communities and Crime Dataset

From Fig. C.28 to Fig. C.32, we show the results of the Communities and Crime dataset. As similar with the results of MNIST dataset, LDA tends to select more features where either of the target cluster or the others have zero or little variance when compared with ccPCA. Also, ccPCA tends to have better separation than PCA.

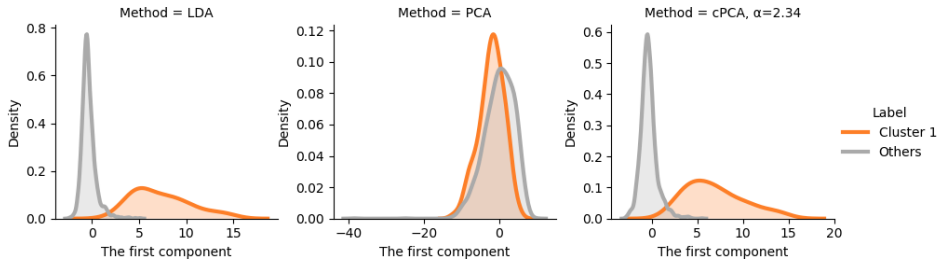


(a) 1D DR result

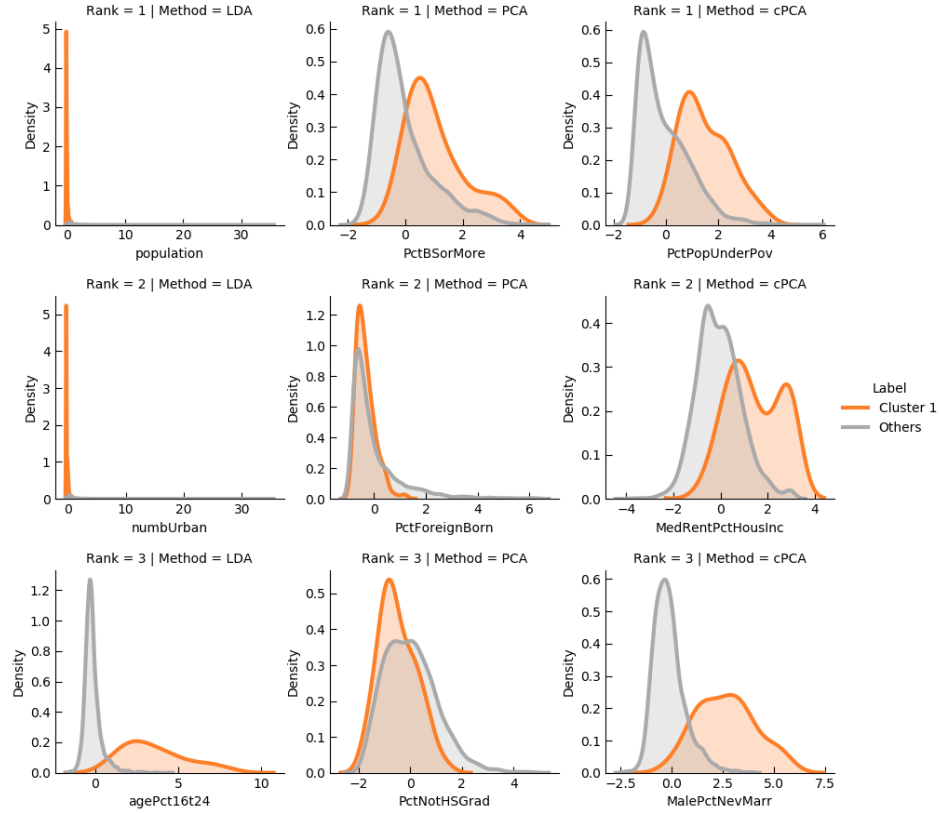


(b) Distributions of the top 3 features' values

Figure C.28: Communities and Crime: Cluster 0

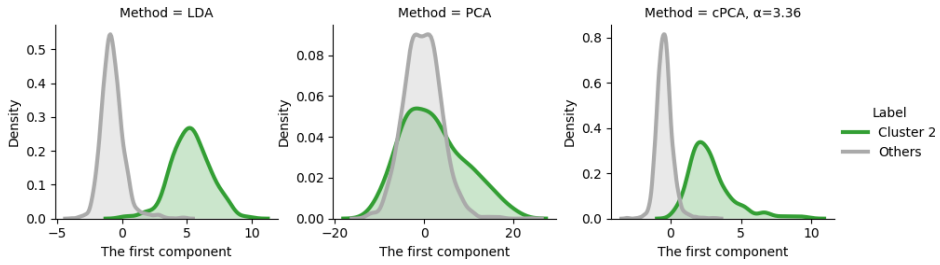


(a) 1D DR result

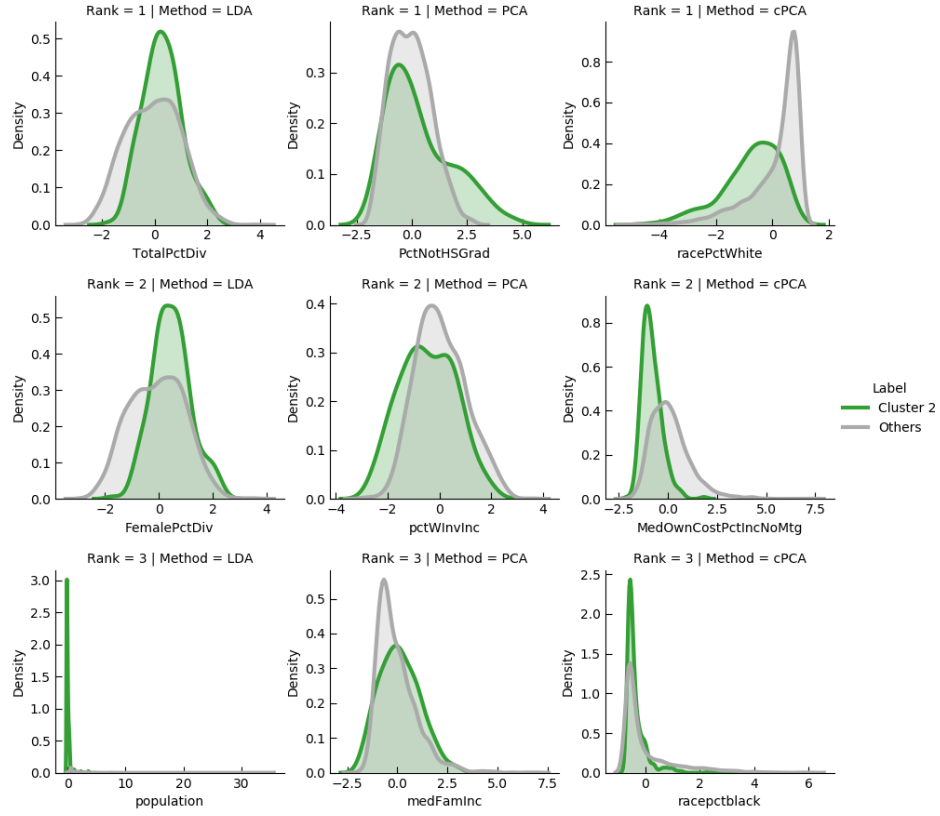


(b) Distributions of the top 3 features' values

Figure C.29: Communities and Crime: Cluster 1

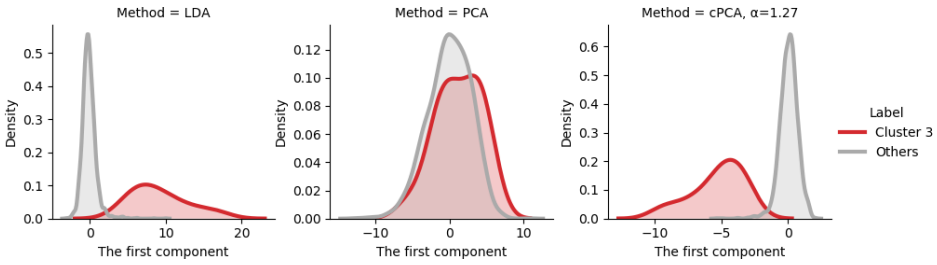


(a) 1D DR result

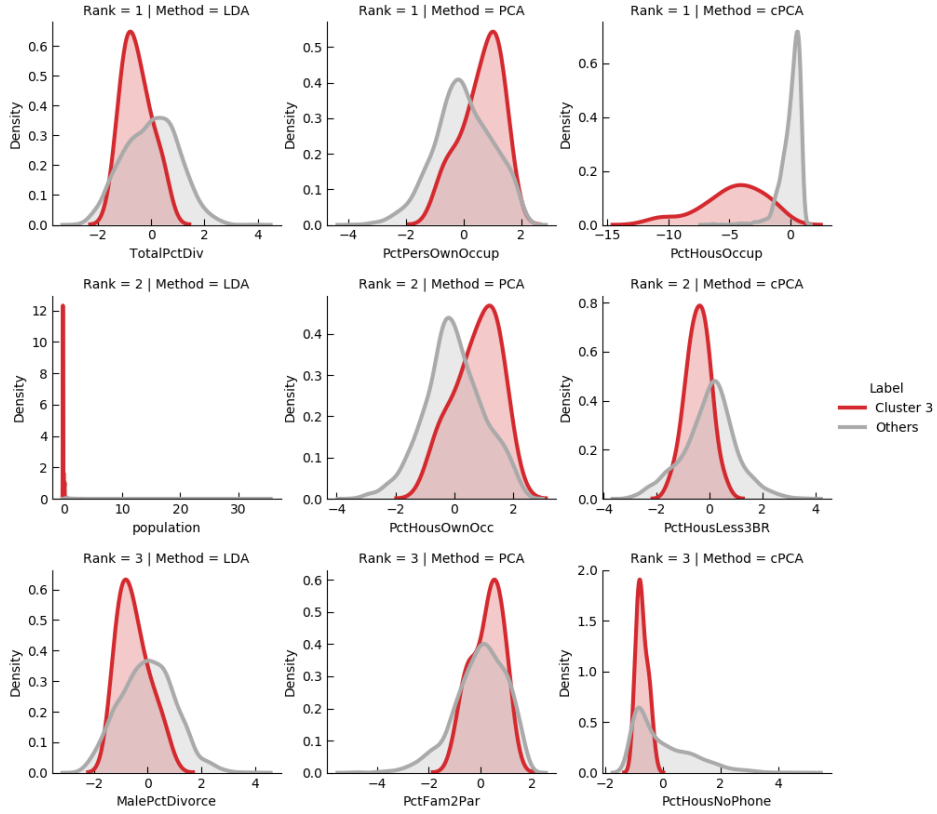


(b) Distributions of the top 3 features' values

Figure C.30: Communities and Crime: Cluster 2

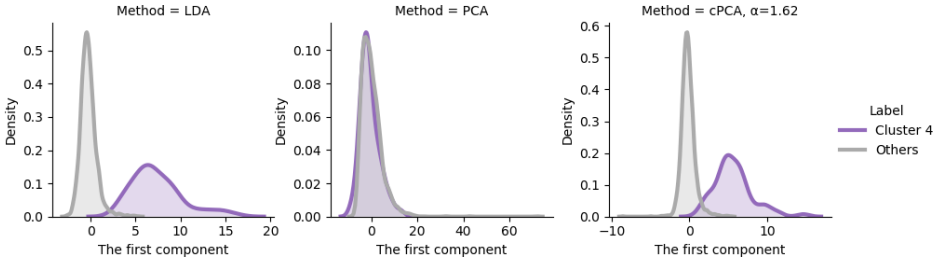


(a) 1D DR result

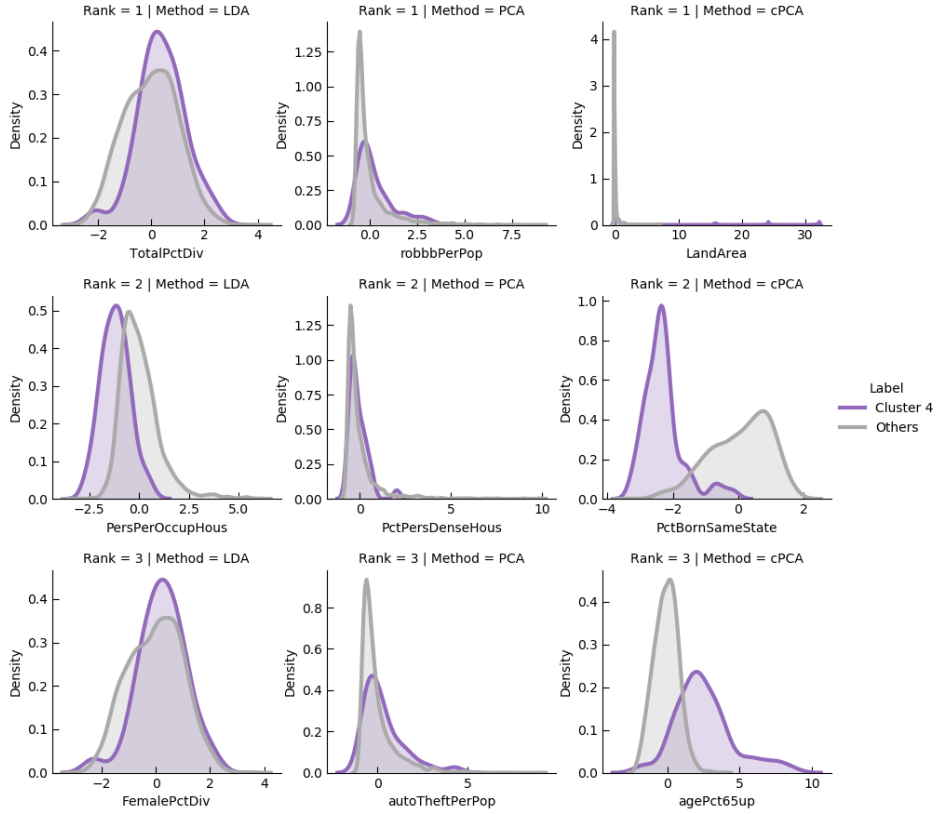


(b) Distributions of the top 3 features' values

Figure C.31: Communities and Crime: Cluster 3



(a) 1D DR result



(b) Distributions of the top 3 features' values

Figure C.32: Communities and Crime: Cluster 4