Analysis Report

1. **PCA Explained Variance Analysis**:
   * The cumulative explained variance graph shows the variance explained by each principal component. Approximately 99% of the variance is achieved around the 175th component, as indicated by the red dashed line. This suggests that a large number of components are needed to capture nearly all variance in the dataset, which may indicate high-dimensional data.

A blue line with a red line

Description automatically generated with medium confidence

1. **PCA Reduced Data Visualization**:
   * The scatter plot of the first two principal components shows the distribution of data in reduced dimensions. This plot allows for visual assessment of patterns or clusters in the data. However, there are no clear separations, which may suggest either a single cluster or overlapping classes in the original data.

A graph with blue dots

Description automatically generated

1. **Bayesian Information Criterion (BIC) Analysis**:
   * BIC scores for different covariance types show that the "diag" covariance type has the lowest BIC score, making it the best option. This finding suggests that the model with diagonal covariance is optimal for capturing the underlying structure in this dataset.

A bar graph with blue squares

Description automatically generated

1. **Optimal Number of Clusters**:
   * The analysis identifies the optimal number of clusters as 1. This result implies that the data does not naturally form multiple clusters, aligning with the single-cluster observation from the PCA visualization.

A graph of a number of clusters

Description automatically generated

1. **Hard Clustering Assignments**:
   * Hard clustering assignments show that all data points are assigned to the same cluster (label "0"). This finding confirms that the data points do not exhibit a strong clustering tendency and are more likely from a homogeneous group.
2. **Soft Clustering Probabilities**:
   * Soft clustering probabilities for each data point indicate a probability of 1.0, which further reinforces the single-cluster finding. This result means that each data point is assigned to the single identified cluster with full confidence.

**Anomaly Detection Report Using Log-Likelihood Scores**

To evaluate whether the model can effectively detect anomalies, we compare the log-likelihood scores for **original (normal)** images and **anomalous** images generated through rotation, flipping, and darkening transformations. The goal is to determine if the model can distinguish between normal data and anomalous data based on the log-likelihood scores produced by the score\_samples() method.

A collage of different faces

Description automatically generated

**Comparison of Log-Likelihood Scores**

Here is a summary of the log-likelihood scores for each category:

* **Original (Normal) Images**:
  + Scores: [−39.71,1.19,−37.17]
  + Average Log-Likelihood: **-25.23**
* **Rotated Images (Anomaly)**:
  + Scores: [−199.51,−120.45,−113.84]
  + Average Log-Likelihood: **-144.60**
  + **Difference from Normal Average**: -119.37
* **Flipped Images (Anomaly)**:
  + Scores: [−66.19,−5.59,−46.40]
  + Average Log-Likelihood: **-39.39**
  + **Difference from Normal Average**: -14.16
* **Darkened Images (Anomaly)**:
  + Scores: [−9.06,−4.29,−9.88]
  + Average Log-Likelihood: **-7.74**
  + **Difference from Normal Average**: +17.49

**Analysis and Interpretation**

1. **Sensitivity to Rotation (Strong Anomaly)**:
   * The rotated images have the **lowest log-likelihood scores** among all transformations, with an average of -144.60, far lower than the original images' average of -25.23. The difference of -119.37 suggests that the model perceives rotated images as highly unlikely given the normal distribution, marking them as clear anomalies.
   * This strong deviation indicates that the model is highly sensitive to changes in spatial orientation, which disrupt the patterns it expects from normal images.
2. **Sensitivity to Flipping (Moderate Anomaly)**:
   * Flipped images also have lower log-likelihood scores than the original, with an average of -39.39. While the deviation is less drastic than with rotated images, it still reflects a significant difference of -14.16 from the normal average, suggesting that the model can detect flipping as an anomaly, though with lower sensitivity than for rotation.
   * This result implies that the model detects spatial changes but is slightly more tolerant of flipping than rotation.
3. **Sensitivity to Darkening (Low Anomaly)**:
   * Darkened images produce scores close to the original images, with an average log-likelihood of -7.74. The difference of +17.49 from the normal average indicates that the model is **relatively tolerant of brightness changes**, perceiving darkened images as more similar to the normal images compared to spatial transformations.
   * This suggests that the model’s primary learned features are based on spatial orientation and structure, rather than brightness.

**Conclusion**

The analysis reveals that the dataset exhibits no significant clustering structure and is best modeled as a single cluster. PCA explains most variance with a high number of components, and transformations such as rotation and flipping significantly affect model fit. The findings are valuable for understanding the dataset’s characteristics and the limitations of the current model in handling spatial transformations.

The model can successfully detect spatial anomalies (rotation and flipping) based on the log-likelihood scores from score\_samples(), as these transformations yield substantially lower scores than the normal images. Rotated images are identified as the most anomalous due to the largest deviation in log-likelihood, followed by flipped images. However, the model is less sensitive to darkened images, indicating that it does not regard brightness variations as significant anomalies.

**Final Evaluation**

* **High Anomaly Detection Sensitivity**: Rotated Images (Score: -119.37 difference from normal)
* **Moderate Anomaly Detection Sensitivity**: Flipped Images (Score: -14.16 difference from normal)
* **Low Anomaly Detection Sensitivity**: Darkened Images (Score: +17.49 difference from normal)

This analysis demonstrates that the model can effectively flag spatially transformed images as anomalies while being more lenient toward brightness changes, highlighting its potential for detecting structural anomalies in image data.

A person's face with different facial expressions

Description automatically generated