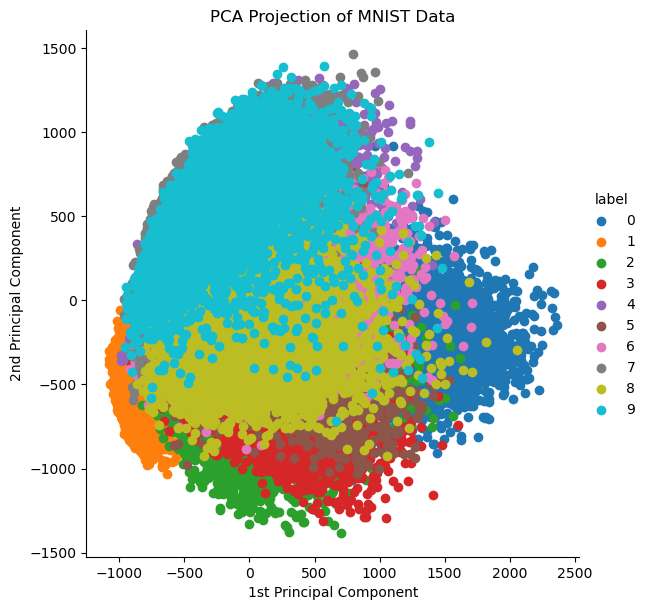
Assignment 1 – Dimensionality Reduction using PCA

**Question 1: MNIST Dataset**

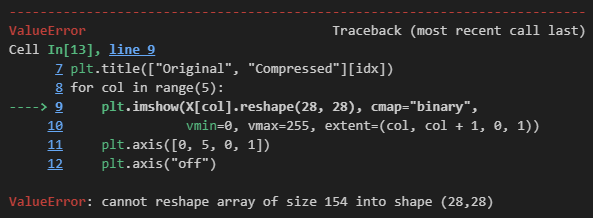
After PCA, the bottom image is the result. Data became dense and computation became faster.

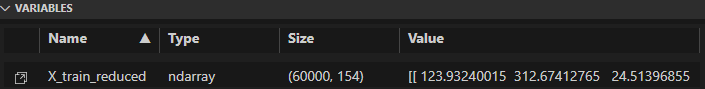


**Challenge Encountered:**

**Plotting X\_train versus X\_train\_reduced**

The shape of X\_train has 784 features as opposed to X\_train\_reduced that has 154. To be able to compare the two, X\_train\_reduced must be returned back to the original feature space by using inverse\_transform.



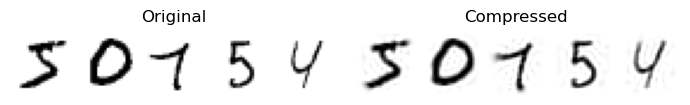
****



****

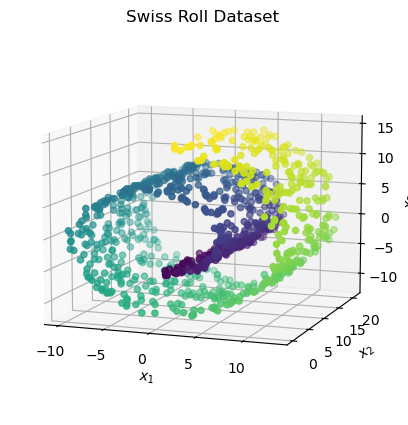
**Result:**

The image below shows the comparison between the original and compressed digits that were generated after using **Incremental PCA and inverse\_transform**.



**Question 2: Swiss Roll Dataset**

The image below is the generated Swiss roll dataset.



The image below resulted from applying linear, rbf and sigmoid kernels to the dataset. However, it is difficult to determine which kernel is best for this model. The best way to decide is use GridSearchCV.

A graph of a number of dots

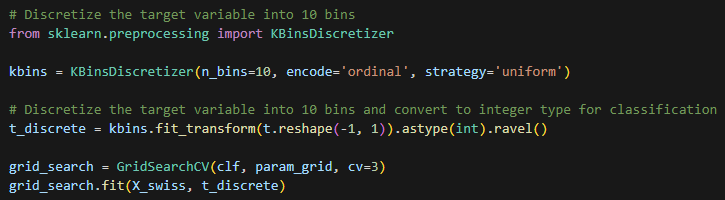
Description automatically generated with medium confidence

**Challenge Encountered:**

**Fitting the grid\_search to X\_swiss together with t calls ValueError**

Since t is an array with continuous values, it cannot be fitted into grid\_search. To solve this, I used feature discretization **KBinsDiscretizer** to change t into integers. Feature discretization decomposes each feature into a set of bins, one-hot encoded and then given to the classifier.

**A computer screen with text

Description automatically generated with medium confidence** ****

**A black and white text

Description automatically generated**

**Result:**

One would think that unrolling the Swiss Roll will result to a better performance for the Logistic Regression. However, upon comparison with the pipeline for choosing the best kernel and gamma values, linear won with 0.03 gamma and a score of 87% whereas unrolling only has 30% accuracy. Also, unrolling the Swiss roll only resulted to more segments.

**A graph with a colorful dotted triangle

Description automatically generated with medium confidence**A graph with a colorful circle

Description automatically generated with medium confidence

**A black background with white numbers

Description automatically generated**