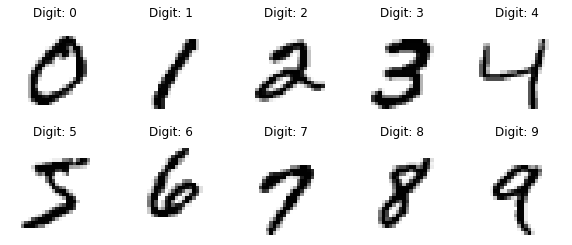
Student name: Artem Kamov

Student ID: 301220613

**Exercise 1**

2.

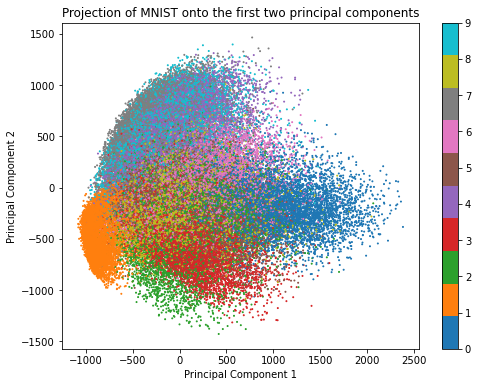


3. Explained variance ratio of 1st compnent: 0.0975 (This value means that the first principal component explains approximately 9.75% of the total variance in the dataset).

Explained variance ratio of 2nd compnent: 0.0716 (This value indicates that the second principal component explains about 7.16% of the total variance).

In total, these two components together explain about 16.9% of the variance in the dataset. It means that most of the variance is spread across other dimensions that are not captured by just these two components.

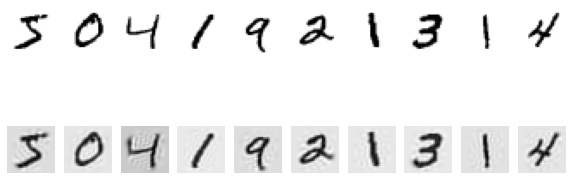
4. Projections of the 1st and 2nd principal component onto a 2D hyperplane



5. Explained variance ratio for IncrementalPCA: 0.95

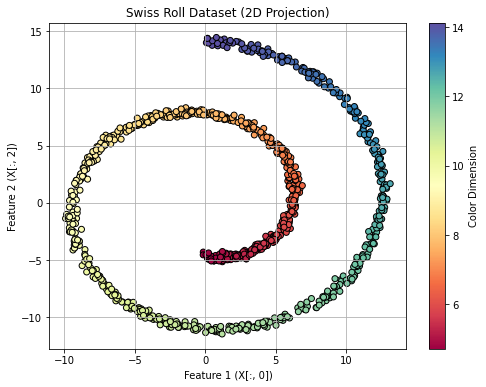
A 95% explained variance ratio means that your reduced dataset (with 154 components) should perform very similarly to the original dataset (with 784 components) for most tasks, while being much more computationally efficient to work with.

6. Original and compressed digits

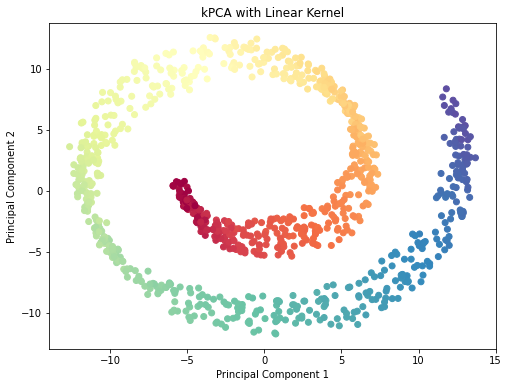


**Exercise 2**

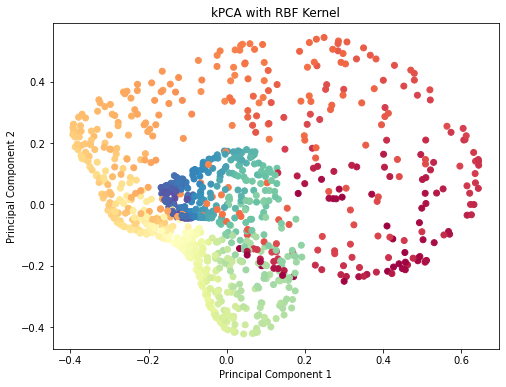
2.



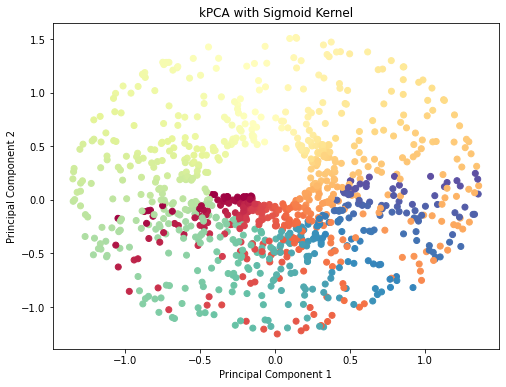
3.



The linear kernel is not effective for non-linear data like the Swiss roll, as it fails to unfold the dataset's complex structure. This is because a linear kernel cannot capture the non-linearity of the Swiss roll's structure which leads to a visualization that retains much of its original shape.



The RBF (Radial Basis Function) kernel significantly transforms the data by mapping it into a higher-dimensional space where non-linear relationships can be captured. This projection appears more scattered indicating that the RBF kernel has effectively "unfolded" the Swiss roll in attempt to separate the data points by their similarities based on the radial basis function.



The Sigmoid kernel also attempts to project the data into a non-linear space, but the result is less effective compared to the RBF kernel. The projection appears somewhat convoluted and does not clearly unfold the Swiss roll as much as the RBF kernel. The central points are tightly clustered, and the outer points remain somewhat intermingled.

To conclude, RBF is often a strong choice for dimensionality reduction using kPCA when working with highly non-linear data

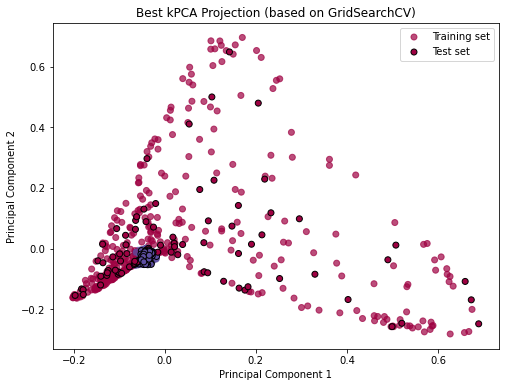
5. Best parameters found: {'kpca\_\_gamma': 0.1, 'kpca\_\_kernel': 'rbf'}

Test accuracy: 0.775

RBF kernel is the best-performing kernel for this classification task. This is typical for datasets with complex, non-linear relationships. The gamma=0.1 parameter controls the shape of the RBF kernel, and the grid search found this value optimal for projecting the data in a way that separates the two classes effectively.

The test accuracy of 77.5% indicates that the combination of kPCA (with the RBF kernel) and Logistic Regression was successful in classifying the two regions of the Swiss roll, though some misclassification may still occur due to the complex overlapping structure of the Swiss roll dataset.

6.



This projection is a result of optimal hyperparameter tuning using GridSearchCV, and it successfully transforms the data into a meaningful lower-dimensional representation that captures the non-linear relationships in the Swiss roll dataset. The plot suggests that the Logistic Regression classifier will likely perform well based on how the data is spread out and how the test set aligns with the training set in this feature space.