**Analysis Report: Non-Linear Dimensionality Reduction with Kernel PCA**

**Author: Ujjwal Poudel**

**1. & 2. Swiss Roll Dataset Generation and Visualization**

A synthetic Swiss roll dataset with 1,000 samples and a noise level of 0.2 was generated to serve as an example of non-linear data. The data, containing X, Y, and Z coordinates along with a color value t, was stored in a Pandas DataFrame named df\_swiss\_Ujjwal.

A 3D scatter plot was created to visualize the dataset's structure. The plot clearly shows a 2D manifold rolled into a 3D space, which presents a challenge for linear dimensionality reduction techniques.

A graph of data visualization

AI-generated content may be incorrect.

**3. & 4. Applying and Comparing Kernel PCA**

KernelPCA was applied to reduce the data from 3 to 2 dimensions using three different kernels to observe their effectiveness at handling the non-linear structure.

**Kernel Configurations:**

Linear: Standard PCA projection.

RBF: gamma=0.002.

Sigmoid: gamma=0.002, coef0=1.

**Results and Comparison:**

The three resulting 2D projections were plotted for comparison. The RBF kernel performed best, successfully "unrolling" the Swiss roll into a flat representation while preserving the neighborhood structure, as indicated by the smooth color gradient. The Linear kernel failed, squashing the roll and keeping the layers overlapped. The Sigmoid kernel produced a similar figure like RBF, however, has bit more overlap than RBF and also has more scattered points.

A diagram of a colorful circle

AI-generated content may be incorrect.

**5. & 6. kPCA for Classification**

A machine learning pipeline was built to see how *KernelPCA* could serve as a preprocessing step for a *LogisticRegression* classifier. The goal was to classify whether a point's unrolled position **t** was greater than **6.9**. *GridSearchCV* was used to find the optimal *KernelPCA* parameters for this task.

**Best Parameters Found:**

Kernel: rbf

Gamma: 0.01

Best Cross-Validation Accuracy: 78%%

**Analysis of Best Configuration:**

The plot of the data transformed by the best kernel shows a clear separation between the two classes. LogisticRegression is a linear model, and it performed best with the RBF kernel because this kernel transformed the intertwined 3D data into a 2D space where the two classes became linearly separable. This enabled the logistic regression model to easily find a decision boundary, resulting in high classification accuracy.

A diagram of a circle with yellow and purple dots

AI-generated content may be incorrect.