# Report on Swiss Roll Dataset Analysis using Kernel PCA and Logistic Regression

## Introduction

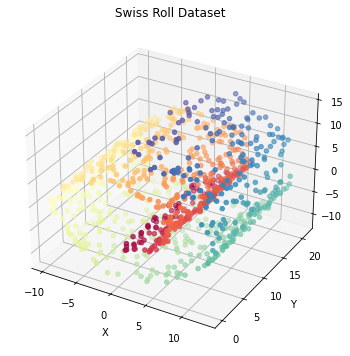
This report presents an analysis of the Swiss Roll dataset using Kernel Principal Component Analysis (kPCA) and Logistic Regression for classification. The analysis includes generating the dataset, applying various kPCA kernels, and utilizing GridSearchCV to optimize classification accuracy.

## A) Swiss Roll Dataset Generation

The Swiss Roll dataset was generated using the `make\_swiss\_roll` function from the `sklearn.datasets` module. The dataset consists of 1000 samples with added noise to simulate real-world data.

## B) Plotting the Swiss Roll Dataset

The generated Swiss Roll dataset was visualized using a 3D scatter plot. The color of the points corresponds to their position along the roll.

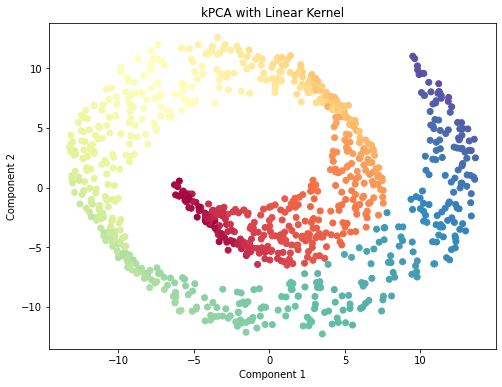
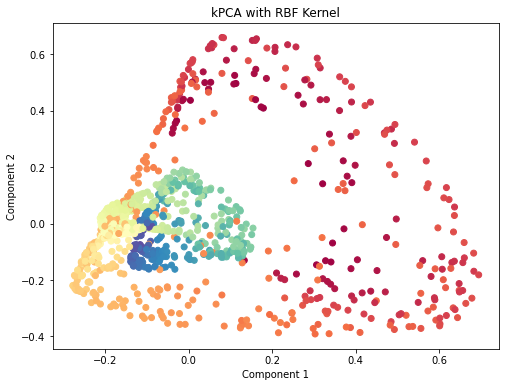
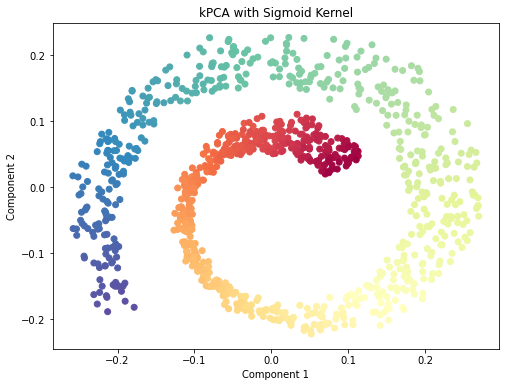


## C) Kernel PCA Application

kPCA was applied using three different kernels:  
1. Linear Kernel  
2. RBF Kernel  
3. Sigmoid Kernel  
  
Each kernel transformation reduced the dimensionality of the dataset to two components.

## D) kPCA Results Visualization

The results from kPCA using different kernels were plotted to compare the effectiveness of each kernel in preserving the structure of the original data.

- Linear Kernel Components:  
  
  
   
  
  
  
  
- RBF Kernel Components:   
  
  
- Sigmoid Kernel Components:  
  
 

## Comparison and Explanation

Linear kernel: Projects the Swiss roll linearly, which may not effectively "unroll" the complex non-linear structure.

RBF kernel: Better at unrolling the Swiss roll, as it can handle non-linearity using a radial basis function, capturing complex structures.

Sigmoid kernel: Does not perform as well as RBF due to its tendency to squish the data, making it difficult to unroll the Swiss roll properly.

## E) Logistic Regression Classification using kPCA

Logistic Regression was applied after kPCA to classify the transformed dataset. GridSearchCV was employed to identify the optimal kernel and gamma value for kPCA to maximize classification accuracy.

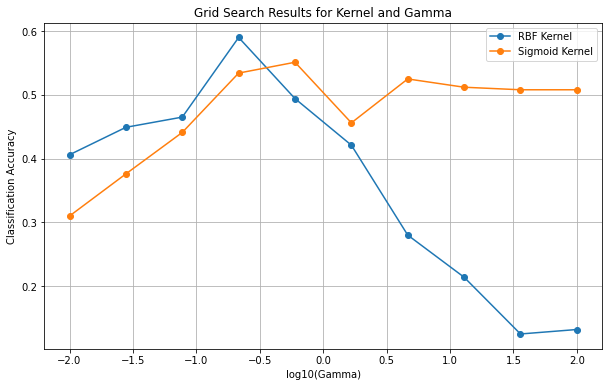
### Results

Best parameters found by GridSearchCV: {'kpca\_\_gamma': 0.21544346900318834, 'kpca\_\_kernel': 'rbf'}

Best classification accuracy: 0.5898443353533174

## F) Visualization of GridSearchCV Results

The results from GridSearchCV were visualized to show the relationship between the gamma value and classification accuracy for both RBF and Sigmoid kernels.

  
The grid search plot shows RBF having peak accuracy in comparison in a certain Gamma parameter. It is also worth noting that a rapid decrease in accuracy in RBF Kernel is seen with higher Gamma values. This is due to overfitting caused by large Gamma parameters.

## Conclusion

The analysis demonstrated the effectiveness of kPCA in transforming the Swiss Roll dataset, allowing for improved classification using Logistic Regression. The comparison of different kernels highlighted the strengths and weaknesses of each method in preserving data structure and enhancing classification accuracy. RBF kernel proved to be most effective with certain parameters in this dataset.  
Further exploration of additional kernels and hyperparameter tuning could enhance model performance. The integration of more advanced classification techniques may also yield better results.