COMP257

Assignment 1: Dimensionality Reduction using PCA

Kai Hong Yeung 301229539

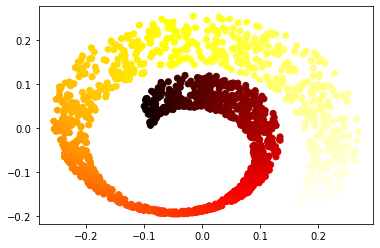
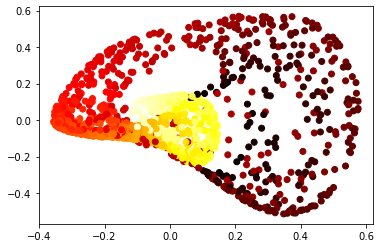
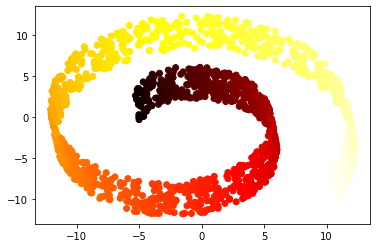
Question 1

Write an analysis report discussing the challenges you confronted and solutions to overcome them, if applicable [15 points]

For the mnist\_784 dataset, there are 784 features and it is hard to choose a good number of dimensions for PCA. However, the sklearn decomposition PCA library already provides the methods to compute the dimensions number. All I need to do is to design the preservation percentage of the portion of the variance.

Question 2

Plot the kPCA results of applying the linear kernel (2 points), a RBF kernel (2 points), and a sigmoid kernel (2 points) from (3). Explain and compare the results in your analysis report [6 points]



For linear and sigmoid kernels, the datasets are projected into 2D without changing a lot. But for rbf kernel, the dataset is projected into a twisted manifold.

In the analysis report discuss challenges you confronted and solutions to overcoming them, if applicable [15 points]

It is hard to design which kernel and parameters are the best for kPCA. The sklearn grid search methodcan help to test it out.