Student Name: Ayman Yasin

Student ID: 301193273

Repository URL: <https://github.com/Ayman850/COMP257.git>

**Written Report**

**Question 1:**

In this question, we loaded the mnist\_784 dataset, which has 70k rows and 784 columns. Next, we plotted the first and second main components into a 1D hyperplane using PCA. Next, we compared the original photos to the compressed images using Incremental PCA to reduce the number of dimensions from 784 to 154.

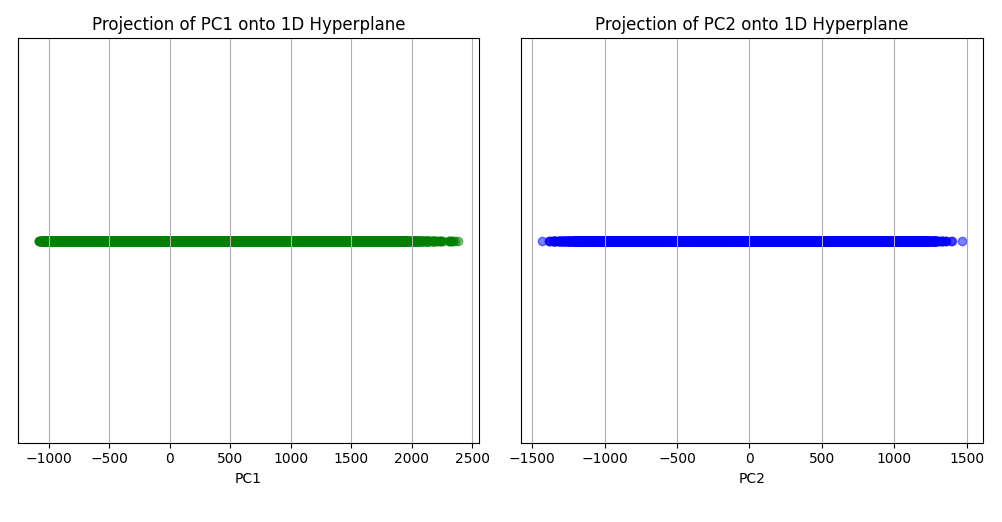
**Images:**

The image below is the plot of each digit from 0 to 9. They’re the first occurrences of each digit in the dataset.

A number in a square

Description automatically generated with medium confidence

The image below is the projections of the first and second principal components onto a 1D hyperplane.



The images below show the difference between the dataset with the original number of dimensions (784) in comparison to the dataset with the compressed number of dimensions (154).

A collage of numbers

Description automatically generated

A number written in white on a black background

Description automatically generated

We can see that Compared to the original, the compressed version has a little bit more blur. In addition, the background appears to be grayer than it was in the original, which was completely black.

**Learning point from question 1:**   
  
Due to the fact that incremental PCA loads the dataset in small chunks rather than all at once, it is better suited for larger datasets.   
  
The compressed images enable us to clearly distinguish the digits despite an approximate 80% reduction in dimension.   
  
Increased Explanation of Deviation By capturing more variance from the original, ratio indicates that compression will be superior.   
  
About 95% of the original mnist 784 could be captured by incremental PCA using 154–D. My discovery was made by adding together the 154 major components' "explained variance ratio."

**Question 2:**

We created a Swiss Roll dataset for this question. Then, we reduced the dimensions from three to two using Kernel PCA using linear, rbf, and sigmoid kernels. Next, we used the Swiss Roll dataset and logistic regression. Since the label was continuous, I just translated so that we may use logistic regression, convert them to integers. Next, we utilized GridSearchCV to determine the optimal kernel and gamma value for Kernel PCA.

The plot of the Swiss Roll dataset that we created is shown in the picture below. It has 1500 samples, and the generation's random state is set to 0.

**Images:**

**A graph of a graph showing a number of dots

Description automatically generated with medium confidence**

**The three charts for the kernel PCA are shown in the image below.**

**A colorful dotted diagram of a brain

Description automatically generated with medium confidence**

The use of the linear kernel produced the plot on the left. Its shape can be seen since it is essentially projecting the 3D Swiss roll onto a 2D surface. It is similar to examining the cross section. We can still see a spiral shape; hence the Swiss roll is not fully unrolled by doing this.

The center plot represents the outcome of applying gamma 0.04 to the rbf kernel. It does a good job of separating the dataset, as we can see. Only a little spiral shape remains, with some parts continuing to overlap.   
The outcome of applying the sigmoid kernel with gamma 0.04 is the plot on the right. As we can see, it still retains a lot of overlaps and the spiral/circular shape from the original Swiss roll.

Since it has the fewest overlaps and eliminates the majority of the original Swiss roll's spiral shape, the plot of the rbf result appears to be the one that unrolls the Swiss roll the best overall. Since it isn't designed to distinguish between non-linear datasets, the linear one appears to be the worse. Non-linear datasets merely cannot be separated using a straight line alone.

The image below shows how gamma affects each kernel type's cross-validation accuracy. It is evident that for both kernels, the gamma value of 0.03 works best.

A graph with a line and a blue line

Description automatically generated

**Learning point from question 2:**Learned more about how the Swiss roll dataset differs in terms of linear, rbf, and sigmoid kernels.   
  
I found that each data point has less of an impact on other data points the higher the gamma value. Stated differently, a low radius of influence corresponds to a high gamma value. Conversely, the radius of influence increases as the gamma value decreases.  
  
Underfitting may result from lower gamma, whereas overfitting may result from higher gamma.   
  
I noticed that the sigmoid kernel may still outperform other kernels in logistic regression, despite its inferiority in terms of visual separation of the Swiss roll dataset.