Homework4

Due: 2019.11.28, 11:59:59

****

For this assignment, please hand in the following two things:

1) A .pdf file contains both your compared results and the explanations of the results.

2) A zip file contains your .py files. Please note that you should write the comment to explain your code.

The ready-made functions, i.e. PCA function in scikit-learn, are not allowed to use in this assignment.

****

Programing exercise:

Load the data from the file **digits-labels.npz** and look for variable **d**. It contains a handwritten digit collection of 28×28 images that is used for handwriting recognition benchmarks. To display the i^{th} data points you can do: imshow(reshape(d[:,i],(28,28),'**F**')). There is also another variable, vector **l**, that contains the digit label corresponding to each column of the matrix d.

1. Select only the columns that correspond to the **digit 5**. Perform PCA and drop the dimensionality down to two dimensions. Plot the projection as a 2D scatter plot like figure1. Make observations on how the shape of the shown digits change as it is distributed across the 2D space.

   Note: You can refer the file show.py to show the 2D scatter plot.

2. Redo this experiment using the embedding from ISOMAP. For the computation of the distance matrix/graph, use only the 6 nearest neighbors. Plot the resulting embedding as before.

3. Redo this experiment using the embedding from Locally linear embedding(LLE). Compare the results of PCA, ISOMAP and LLE.

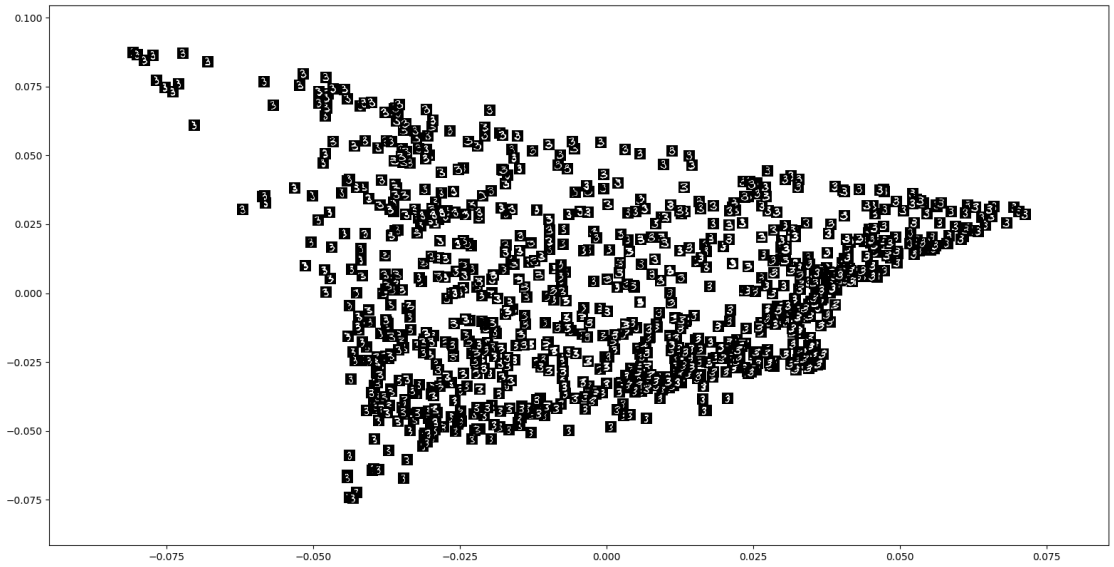

Figure1, LLE result on digit 3.