Practical Lecture 2

Data visualisation: Projection models (Manifold learning)

Part I. Swiss-roll dataset

1. Generate the swis-roll dataset with 2000 points using the function datasets.make_swiss_roll

2. Apply the PCA and plot the data.

3. Apply LLE (Local Linear Embedding) with 5 neighbours (manifold.locally_linear_embedding) by printing the error. Change the number of neighbours from 2 to 15 and plot the error line. Which is the best number of neighbours ?

4. Use Multi Dimensional Scaling with manifold.MDS and visualize the dataset in 2 dimension.

5. Apply t-SNE model to the same dataset with manifold.TSNE.
Visualize the dataset.

6. Conclude and give the details of the best model.

Part II. Digit dataset

Import the digit dataset containing only 6 classes :
digits = datasets.load_digits(n_class=6)

Follow the tutorial :
https://scikit-learn.org/dev/auto_examples/manifold/plot_lle_digits.html

1. Analyse the results by explaining the models
2. Use a classification model (Decision Tree for example) on all the projections and compute the errors
3. Compute the quality of models and highlight the gain.