DA5401 Assignment 9

You might be aware that CoDS-COMAD 2025, the prestigious international data science conference is happening by the end of the year. As a part of the conference, there is an associated data challenge to predict the key attributes from product images. Given an image of a dress item (in 5 categories: Men Tshirts,Sarees,Kurtis,Women Tshirts,Women Tops & Tunics), the task is to predict the attributes of the image such as color, sleeve_styling, transparency, fit_shape, pattern, length, etc. Each category of the dress item may have a different number of dress attributes. To download the data, you may have to fill a form to get access to the competition page.

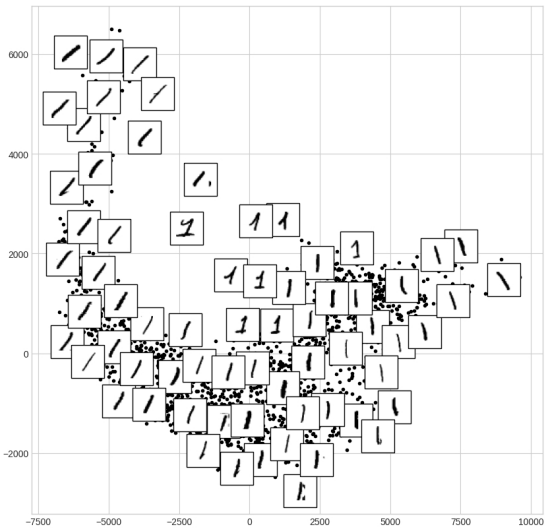
We are going to use our knowledge from manifold learning and dimension reduction lectures to visualize the dataset and discover interesting patterns and their association with product attributes.

Task #1 [15 points]

The challenge dataset contains ~70k training image in 5 categories with the respective attributes. For each category, pick any two attributes of your choice (say color or length or pattern, or any) to form baskets. Each basket is a <category, attribute> tuple. You will create two tuples per category, so in total you should have 10 baskets. Draw 100 samples from each basket. If your basket does not have 100 samples, reconfigure your basket by changing the attribute.

Task #2 [20 points]

Create the visualization like below (which we reviewed in the class) for each basket. You should use Isomap and tSNE with two components, which would represent the intrinsic dimensions of the manifold on which the dataset resides. You will have 10 visuals using Isomap and another 10 visuals vis tSNE.



Task #3 [15 points]

Now comes the interesting part. Recognize the patterns and figure out a name for the components your manifold learning methods have discovered. You should also reason your choice of the name to the discovered manifold dimension.