Final Project - Testing Procedure

Wednesday, April 27, 2022 6:21 PM

* Importing Data

	Train	Test	Dimensions		Classes	
✓ MNIST	60k	10k	784	28×28	10	(0-9)
✓ Fashion MNIST	60k	10k	784	28×28	10	clothes, tshirts etc.
✓ Olivetti Faces	400		4,096	64×64	40	individuals
*✓ COIL-20	1440		16,384	128×128	20	objects
Word Feature						
X Netflix						

FOR MNIST and Fashion MNIST, We considered 2% of training data i.e (2% of 60k = 1.2k)

* Fashion MNIST data used instead of COIL-20 as it was very computationally heavy

* Using PCA to reduce High-Dimension to 30-D [Mentioned in Paper]

This eases computation
All 3 datasets are reduced to 30-D and their pictures have been shown in code
→ MNIST, Fashion MNIST, Olivetti Faces

Data normalized to [0,1]

* Dynamic Visualization of SNE and tSNE to show how they both perform in data separation

Diff.

SNE	tSNE
uses softmax-related gaussian distribution	Uses t-distribution where closer datapoints are brought more closer and further points are more further apart

Uses slightly differ Cost function

* Comparison Graph

tSNE Vs

SNE	perplexity = 40	hyperparameter
Isomap	k = 12	
LLE	k = 12	
UMAP	k = 12	
Sammon Mapping	None	

Comparison done on all 3 datasets

MNIST: UMAP > tSNE > LLE > Isomap > SNE

Fashion MNIST: UMAP > tSNE > SNE* = LLE* > Isomap

Olivetti Faces: Inconclusive

* Close results so both are equally acceptable

Why UMAP > tSNE

Primary difference is in cost function

UMAP implements a cost function which tends to separate on global structure, meaning it even considers same classes that are further apart and tries to separate them from other classes and form a cluster

Unlike

t-SNE has a cost function which tends to separate on local structure, meaning it cannot consider same classes that are are further apart. It will only consider closer neighbouring datapoints and tries to separate them