Phase Transition in TSNE : Critical sample size for learning

Laura Bonde Holst S173953
Pranjal Garg s210242
Virgile Ulrik Blanchet-Møhl s163927

Who does what

____ Virgile

Introduction

Models

_____ Laura

Experiment

Results

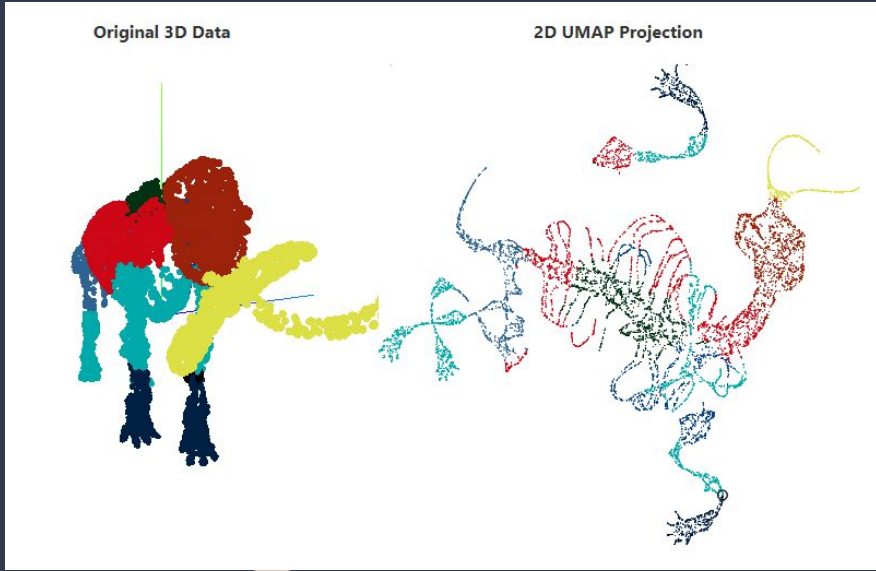
— Pranjal

Discussion

Future Work

Conclusion

Introduction



- Dimensionality reduction is an important tool for visualization of high-dimensional data
- Multiple embedding algorithms have been proposed and see use
- Our work studies the impact on learning of:
  - Missing data
  - Noise level
  - Class imbalance

DR Models

PCA	t-SNE	UMAP	Trimap
Linear dimensionality reduction based on finding direction that maximize variance	Stochastic neighbor embedding using t-distribution	Produces high-dimensional graph of data optimized into a low-dimensional representation	Constrained triplet neighbor embedding using added random neighbors
PCA performance is depending on noise and sample size	Focuses on local structure by weighing nearest neighbors increasingly	Decreasing likelihood of neighbor connection with distance to preserve local and global structure	Preserves local structure with constrained neighbors

PCA

t-SNE

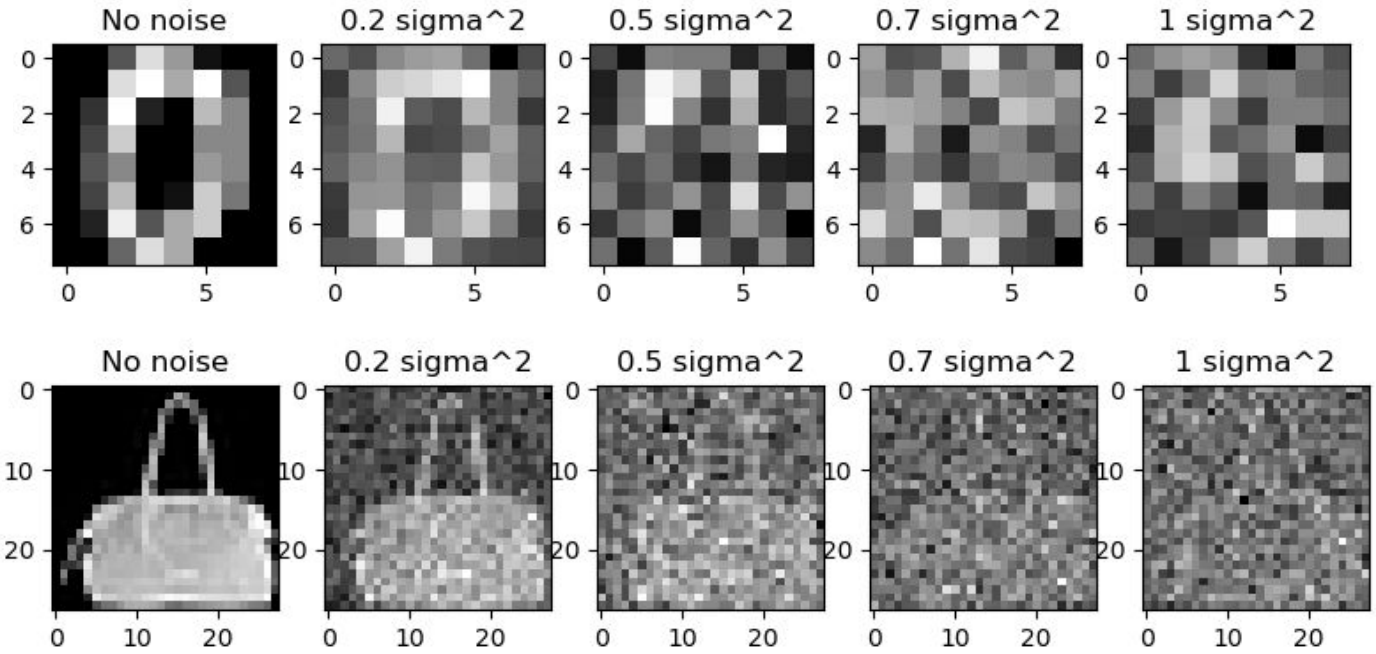
UMAP

TriMap

Experiment

DR Techniques	*Datasets*	*Noise levels*	*Class distribution*	*Sample size*
PCA t-SNE UMAP TriMap	MNIST (8x8) (0+1) Fashion-MNIST (28x28) (Trousers + sneakers)	$\mu = 0, \sigma^2 = 0$ $\mu = 0, \sigma^2 = 0.2$ $\mu = 0, \sigma^2 = 0.5$ $\mu = 0, \sigma^2 = 0.7$ $\mu = 0, \sigma^2 = 1$	Stratified at 50/50 Stratified at 25/75	4 datapoints . . . 90 datapoints

Experiment: Noise levels



Experiment

DR Techniques	*Datasets*	*Noise levels*	*Class distribution*	*Sample size*
PCA t-SNE UMAP TriMap	MNIST (8x8) Fashion-MNIST (28x28)	$\mu = 0, \sigma^2 = 0$ $\mu = 0, \sigma^2 = 0.2$ $\mu = 0, \sigma^2 = 0.5$ $\mu = 0, \sigma^2 = 0.7$ $\mu = 0, \sigma^2 = 1$	Stratified at 50/50 Stratified at 25/75	4 datapoints . . . 90 datapoints

Experiment: Accuracy

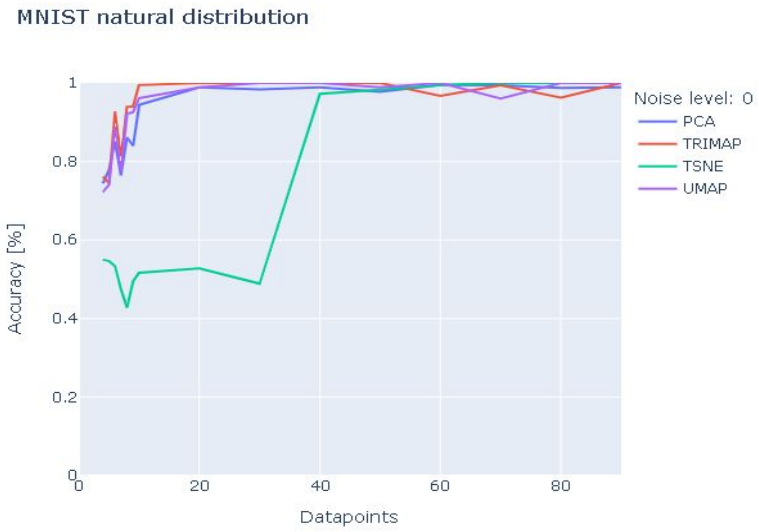
1. K-Means clustering
2. Calculate accuracy

$$accuracy = 1 - \frac{number\ of\ false\ classifications}{total\ number\ of\ classifications}$$

Experiment: Experiment flow

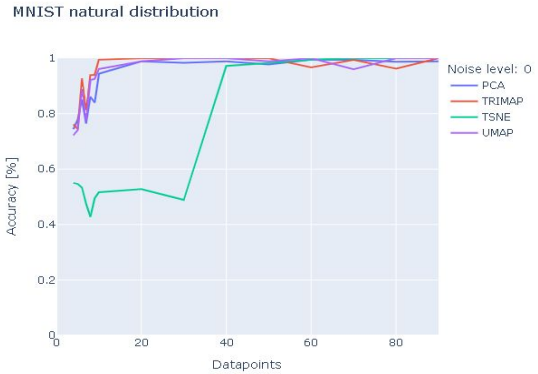
Results:

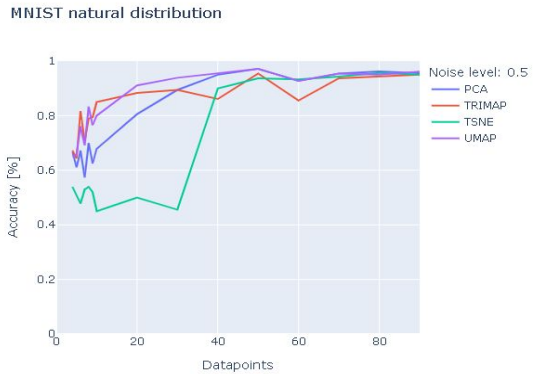
Dataset	*Noise levels*	*Class distribution*
MNIST	$\sigma^2=0$	50/50

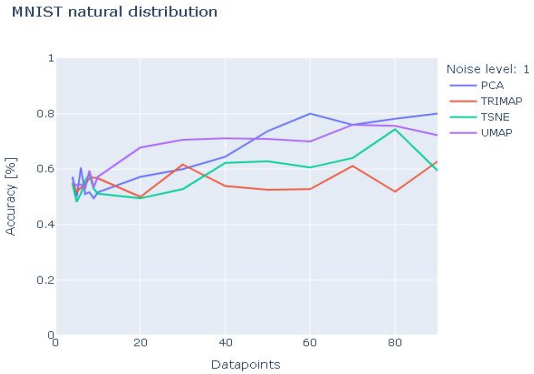


Results:

Dataset	*Noise levels*	*Class distribution*
MNIST	Varying	50/50

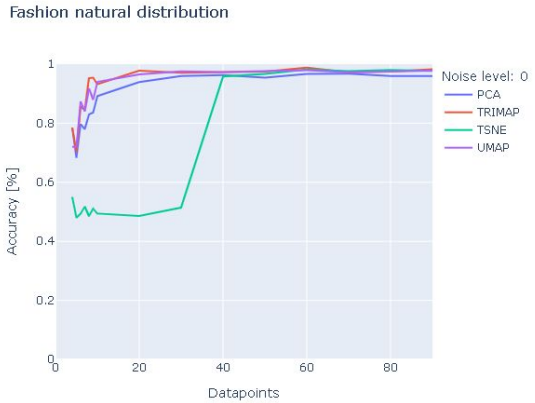


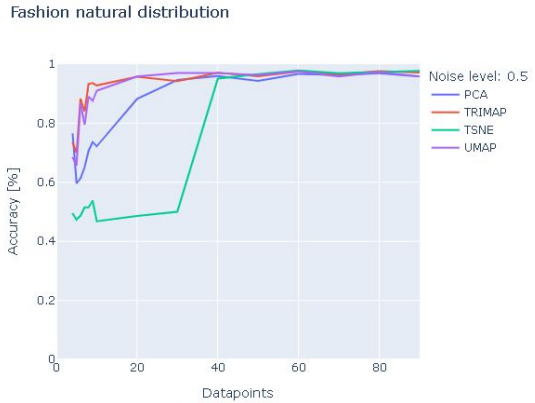


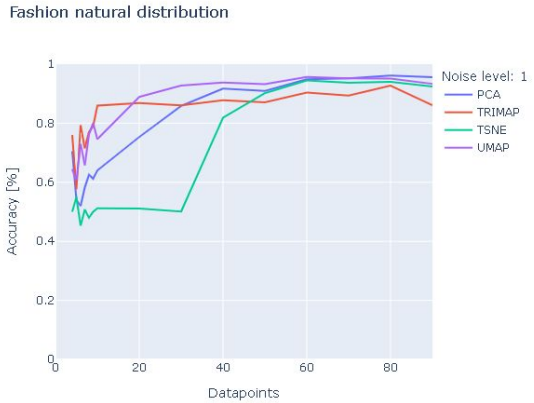


Results:

Dataset	*Noise levels*	*Class distribution*
Fashion-MNIST	Varying	50/50

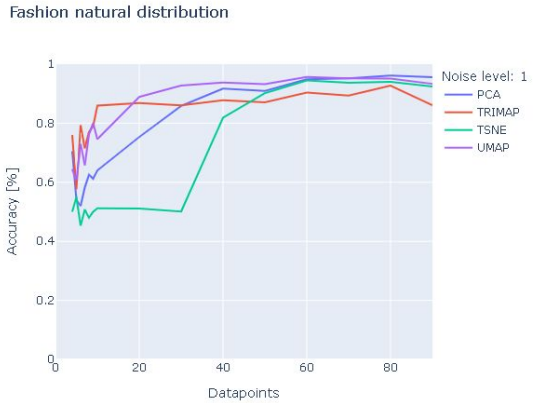


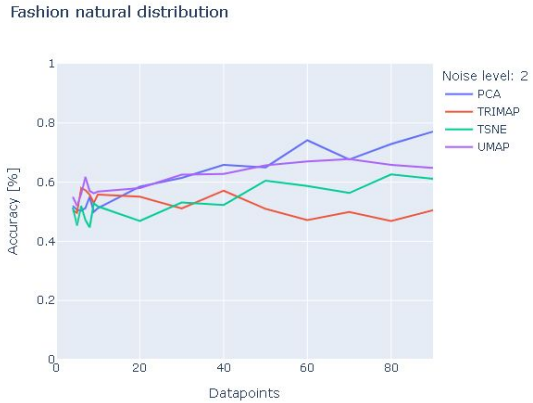


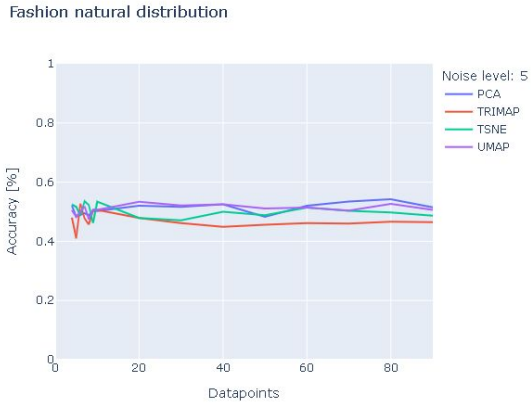


Results:

Dataset	*Noise levels*	*Class distribution*
Fashion-MNIST	Varying (high)	50/50

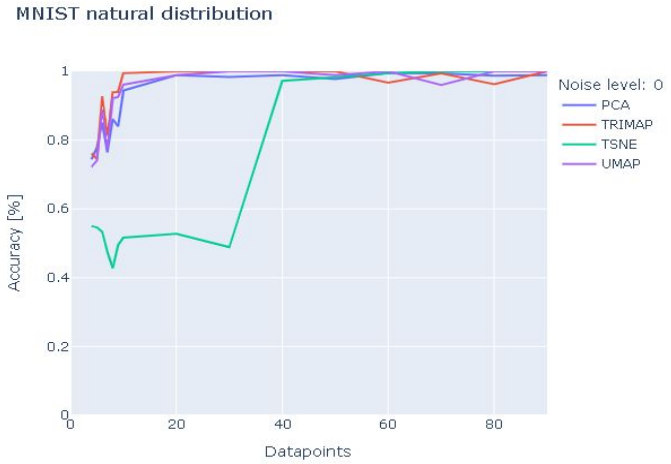


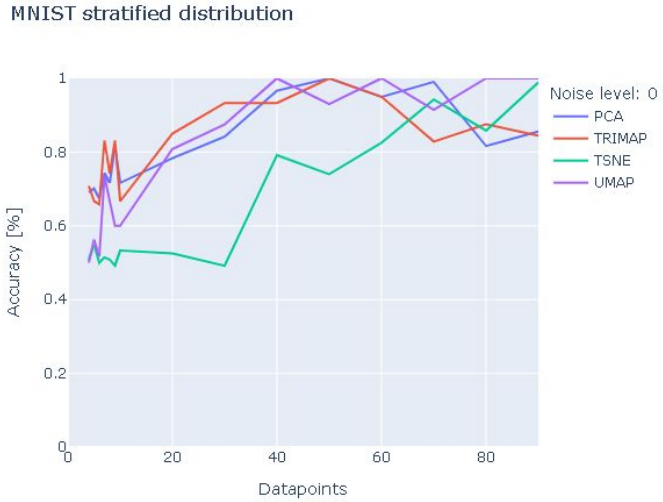




Results:

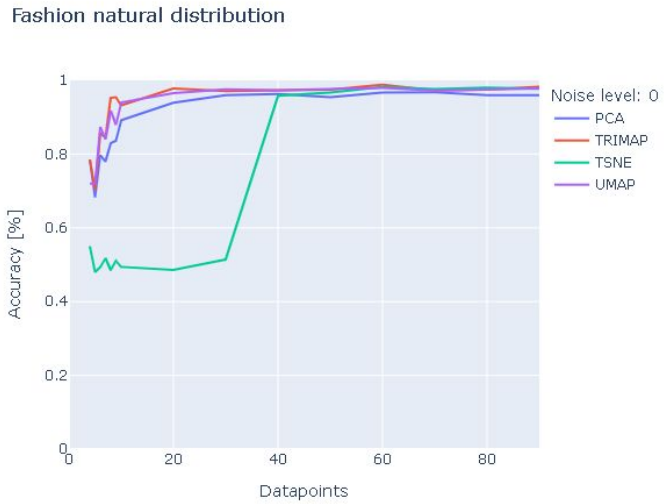
Dataset	*Noise levels*	*Class distribution*
MNIST	$\sigma^2=0$	Varying

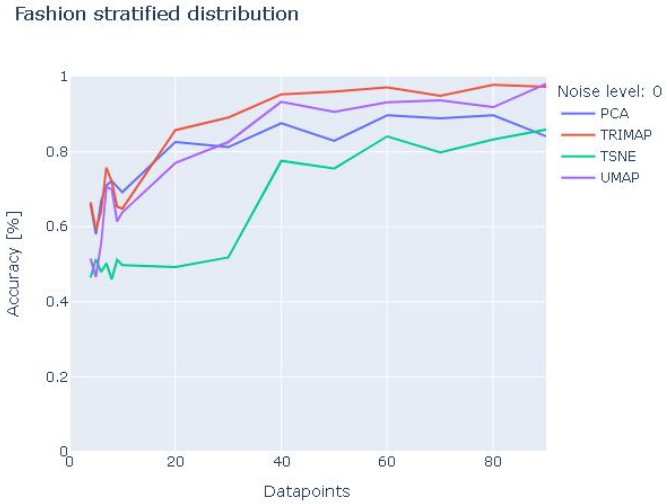




Results:

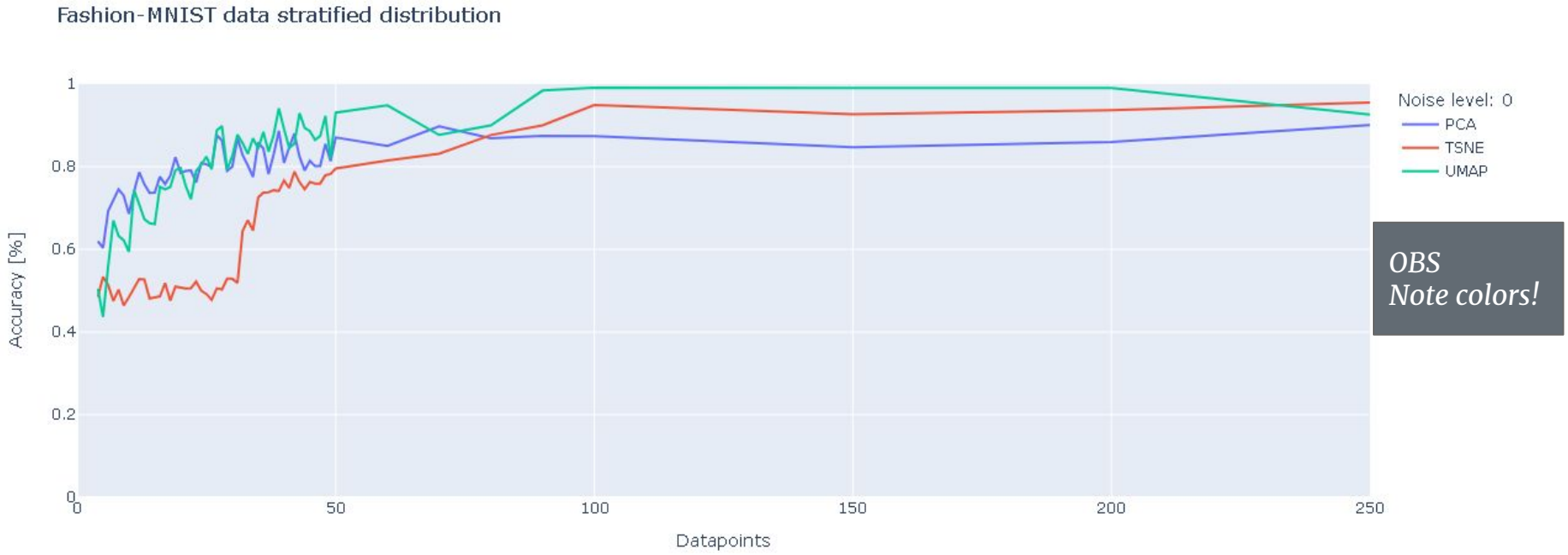
Dataset	*Noise levels*	*Class distribution*
Fashion-MNIST	$\sigma^2=0$	Varying





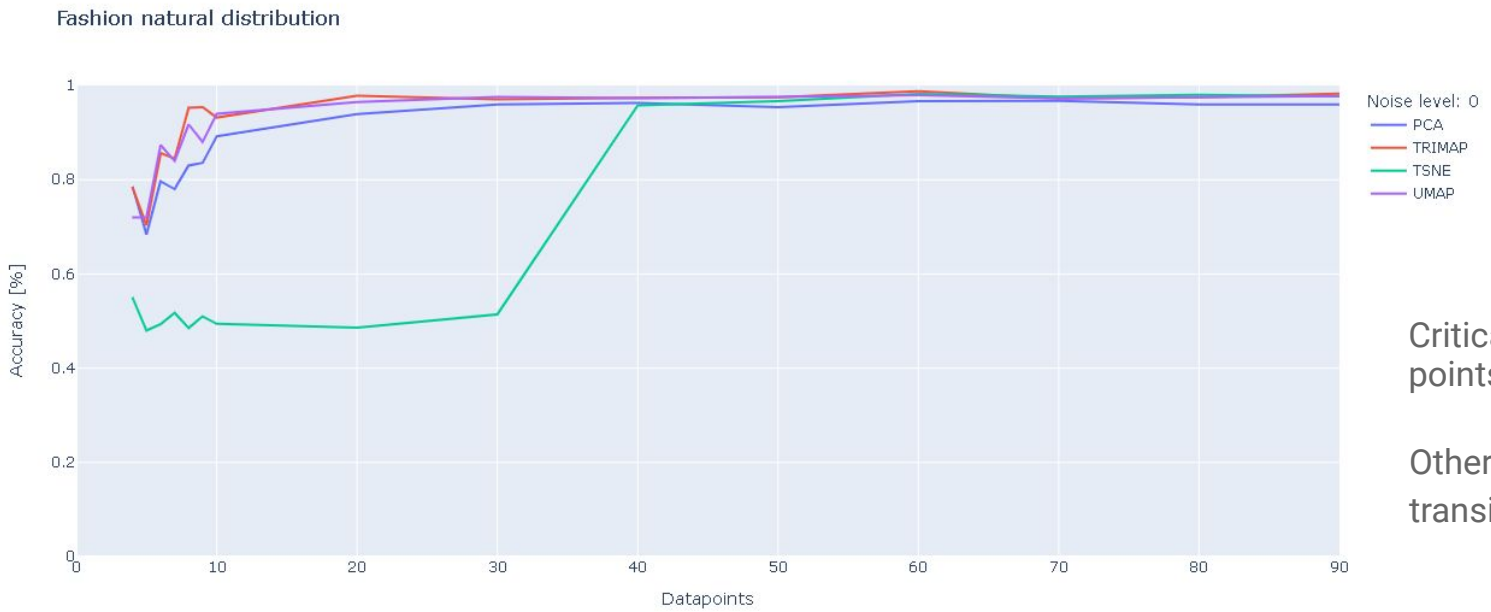
Results:

Dataset	*Noise levels*	*Class distribution*
Fashion-MNIST	$\sigma^2=0$	25/75



OBS
Note colors!

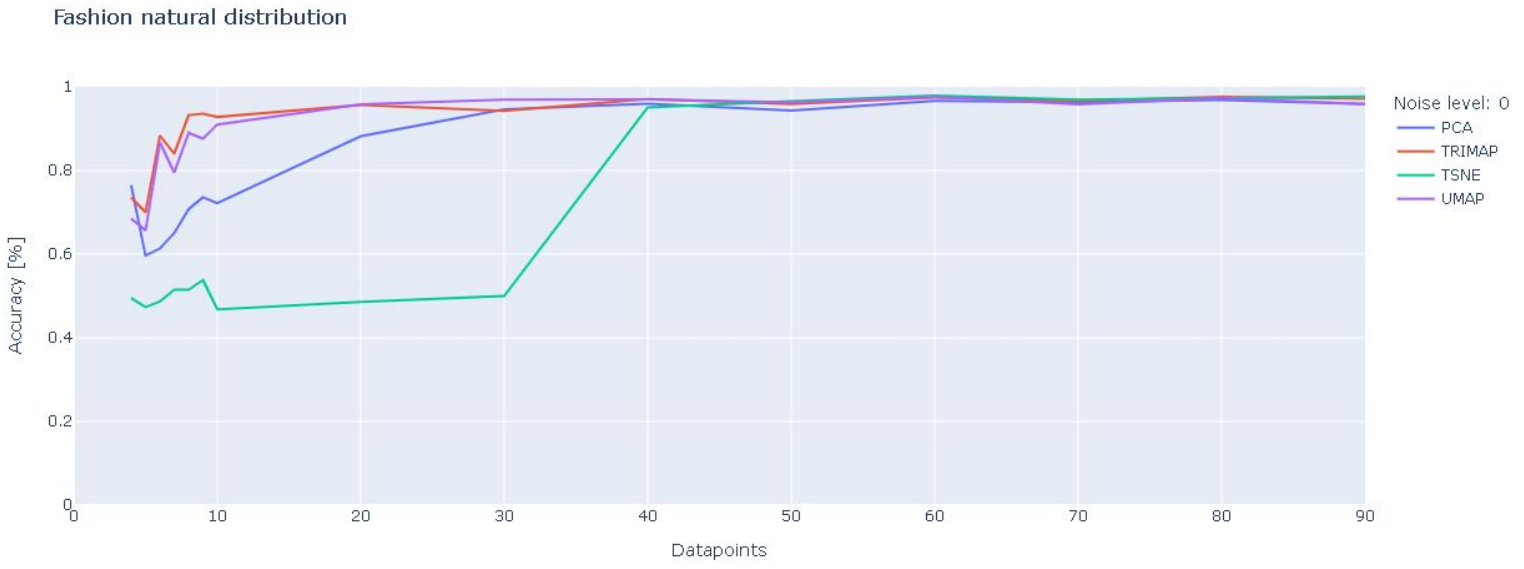
Discussion : Phase transition in t-SNE

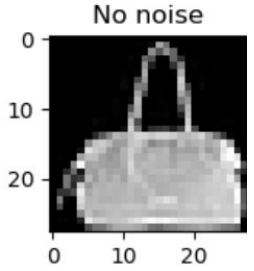


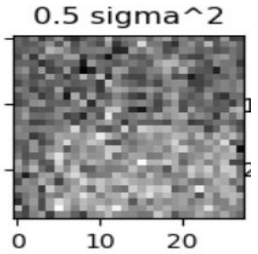
Critical Threshold over 30 points for learning in t-SNE.

Others show slight phase transition between 4 and 10.

Discussion : Effects of noise





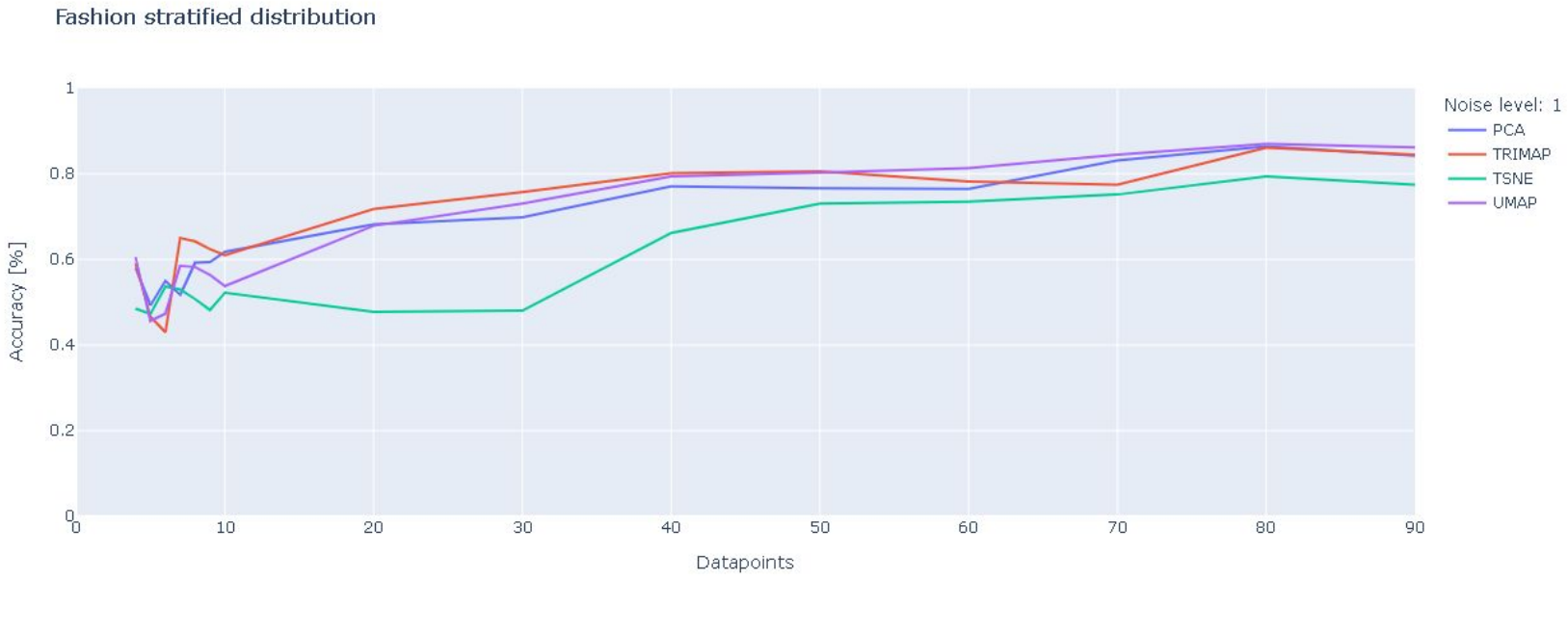


t-SNE's Magic number 40!

PCA shows gradual learning till 40

UMAP , Trimap still at 10.

Discussion : More noise

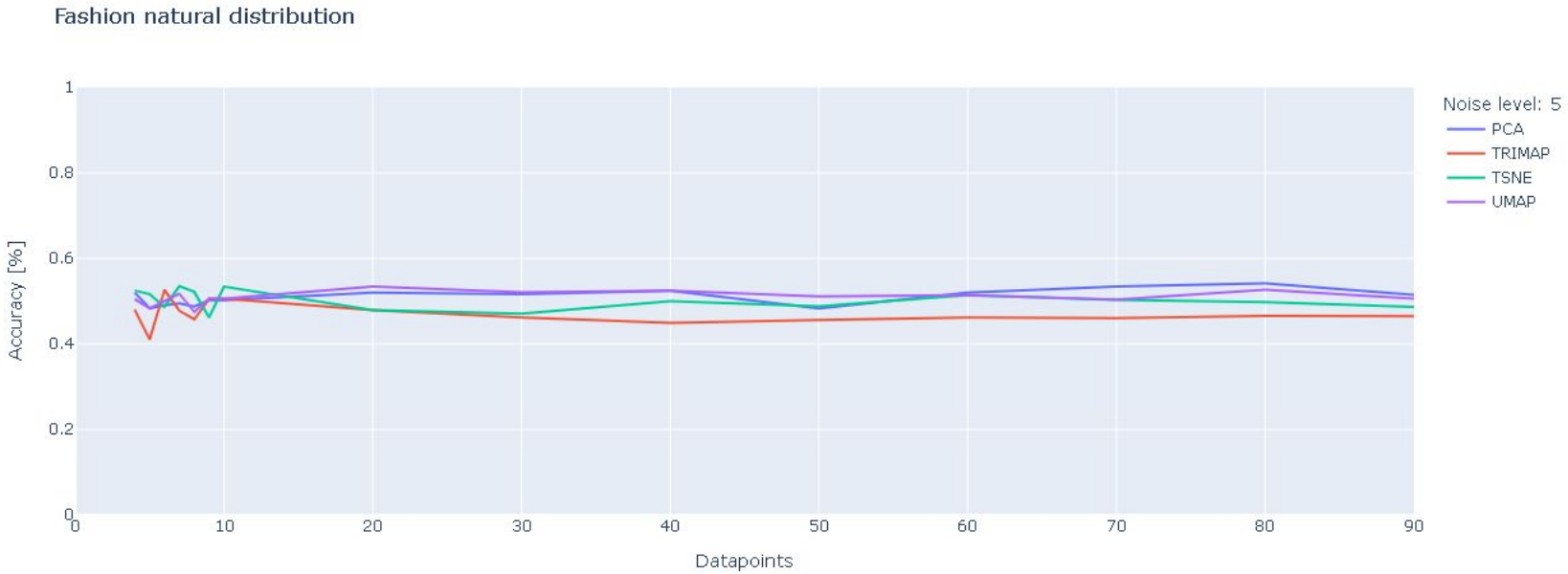


Others showing steeper learning till 40 points too

Slope of learning curve reduced

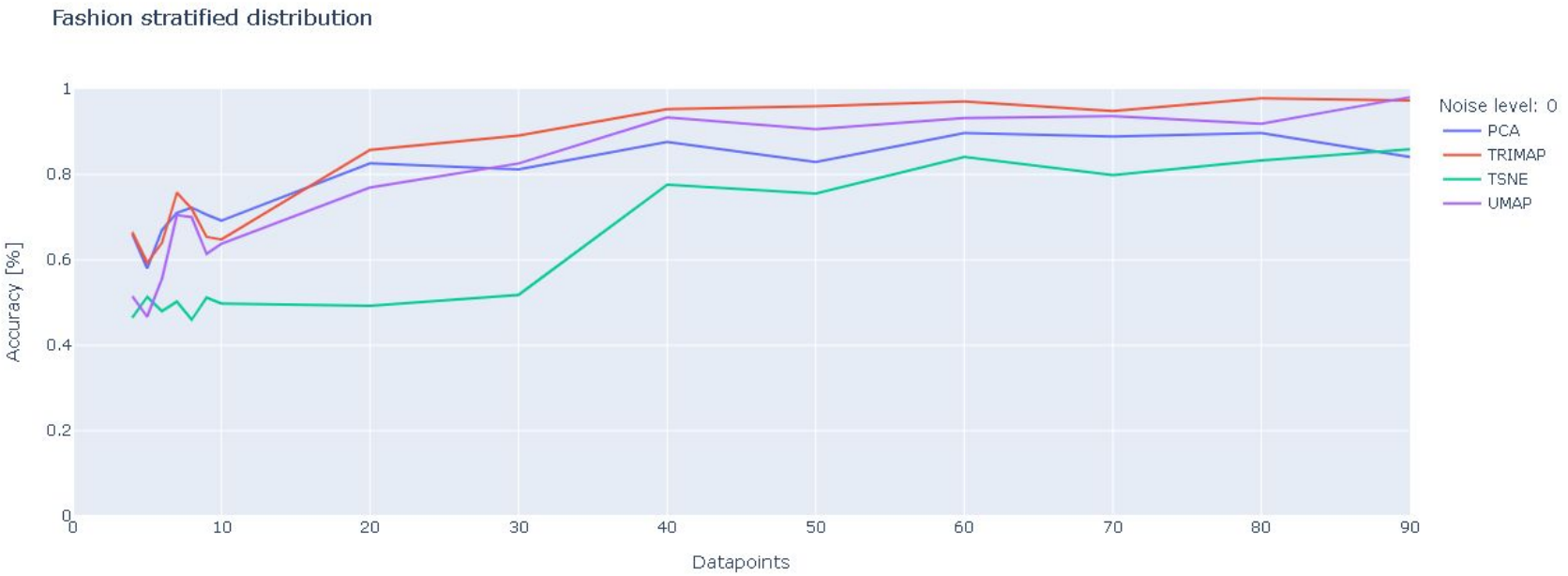
Decreased accuracy

Discussion : Very noisy data



All signal lost

Discussion : Imbalanced classes

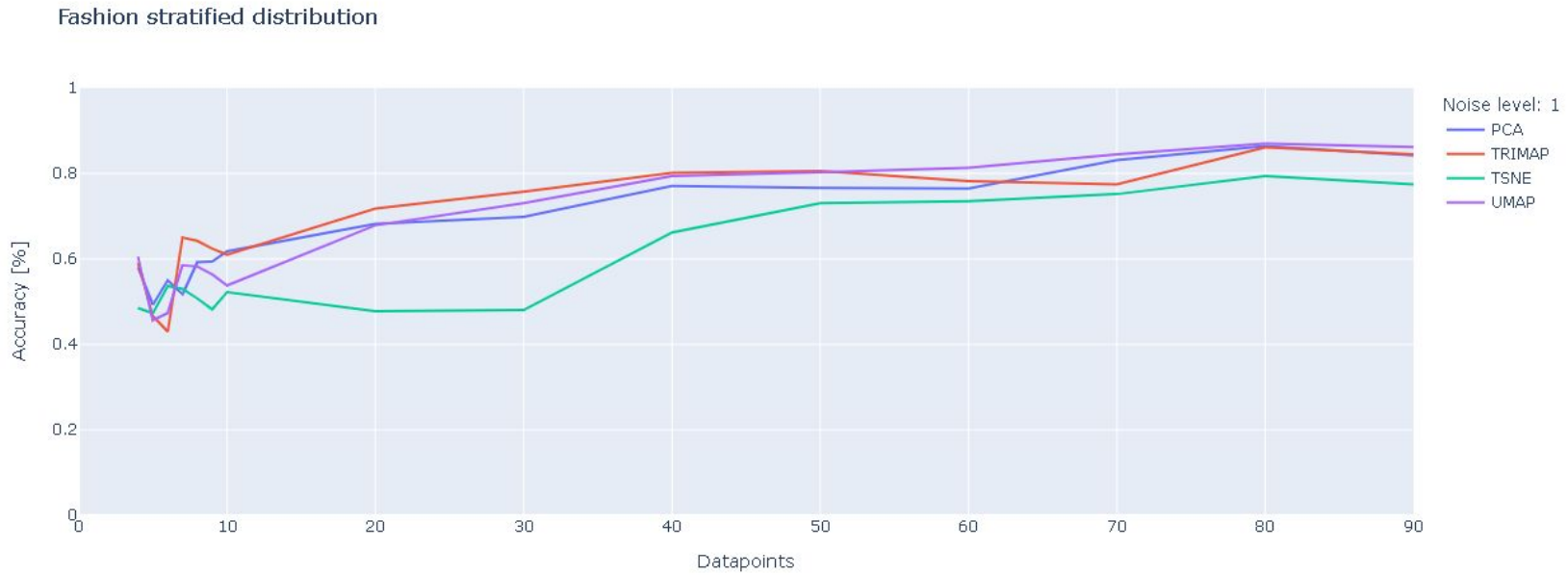


Similar effects to noise

More profound effect on accuracy

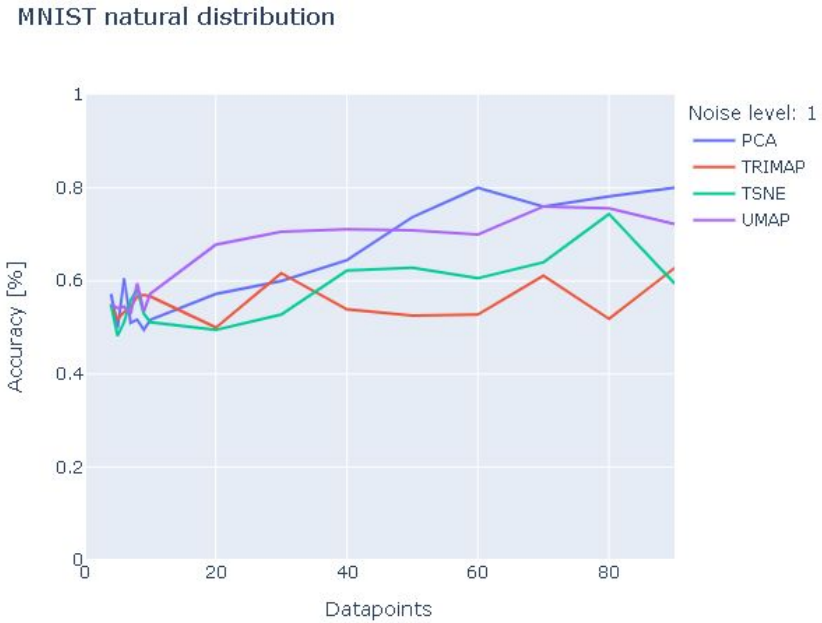
More consistent learning for other techniques till 40?

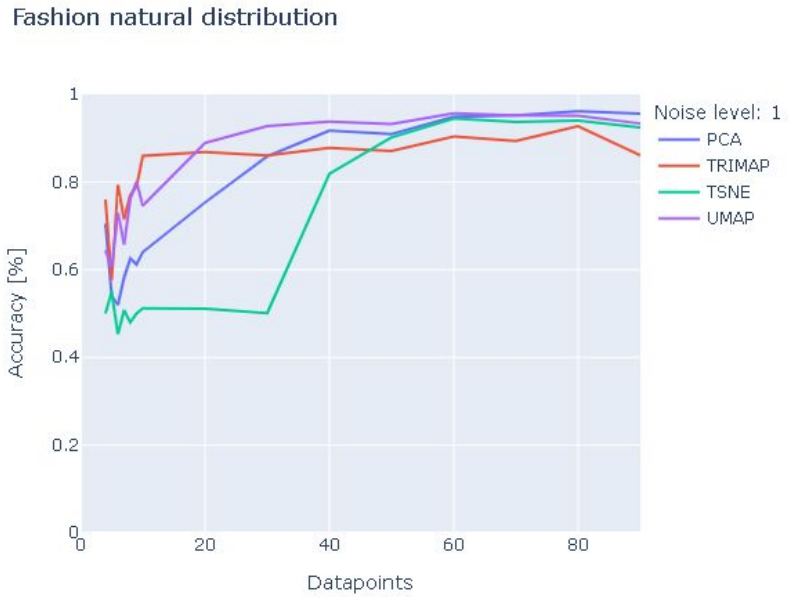
Discussion : Class imbalance with noise



Higher learning rates till 40 for others and phase transition after 30 in t-SNE

Discussion : Dimensionality effects





- Higher dimensional dataset more robust to noise but do degrade with even higher noise levels
- Need Signal to Noise ratio measure to better understand effects of dimensions.

Discussion Summary

- In our experiments, we consistently see a phase transition in TSNE between 30 and 40 data points.
- Only with strong noise signals the phase transition disappears which is due to the original signal being nullified
- The other models perform decently from beginning with step wise improvement in their learnings.
- With stratification, you do see slight phase transitions in the other models which could potentially indicate a need of minimum data points for their learnings as well.

Future Work

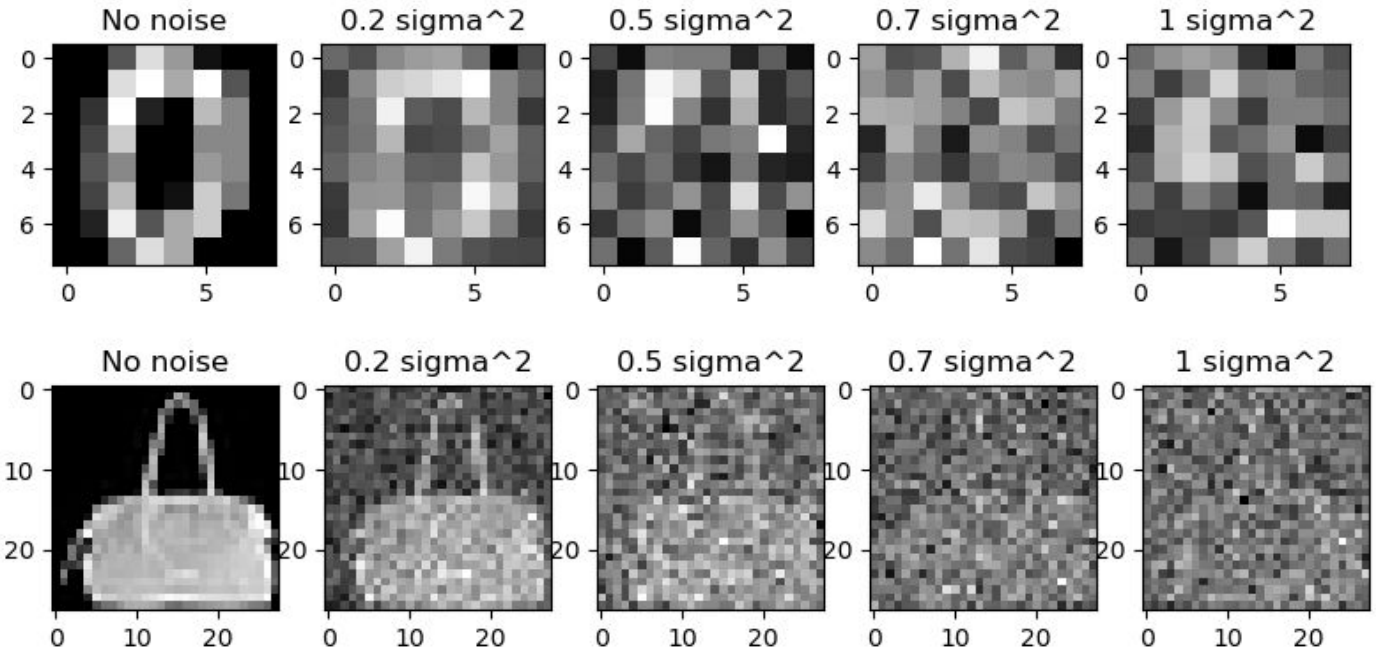
- Simulated data with extremely high and low dimensions.
- Signal-to-noise ratio.
- Test on other DR techniques like LargeVis, Kernel PCA , Laplacian Eigenfolds. and PAC-MAP
- Missing data.

Conclusions

- A critical number of points are needed for learning to start from t-SNE.
- No other DR methods that we tested showed such a phase transition.
- Learning speed is affected by noise in the signal and class imbalance.
- A constant phase transition around 30-40 for t-SNE in our experiments.
- Future work needed to reaffirm experiment results without using k-means.

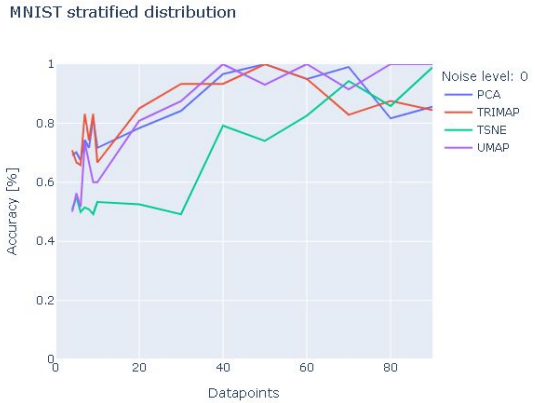
Appendix

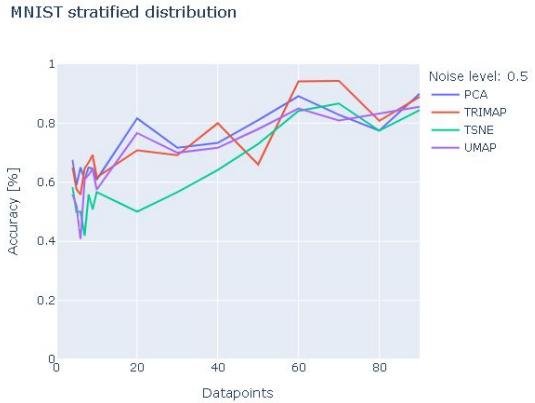
Appendix: Noise levels

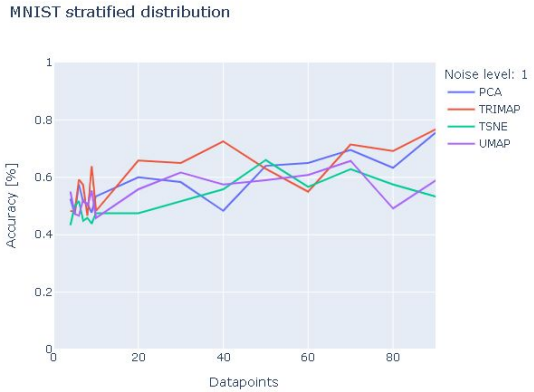


Appendix

Dataset	Noise levels	Class distribution
MNIST	Varying	25/75







Appendix

Dataset	Noise levels	Class distribution
Fashion-MNIST	Varying	25/75