Optimisation Methods:

Task 1 (Analysis of speed and accuracy of the optimization of defined functions by the methods):

	No. of iterations needed	Error
Rosenbrock	NelderMead :55 Simulated Annealing:1000 CMA-ES:94	NelderMead :$3.7*10^{-10}$ Simulated Annealing:$7.01*10^{-12}$ CMA-ES: $1.44*10^{-14}$
Rastigrin	NelderMead :31 Simulated Annealing:1000 CMA-ES:79	NelderMead : $1.932*10^{-9}$ Simulated Annealing: $7.105*10^{-15}$ CMA-ES: $1.78*10^{-15}$
Ackley	NelderMead :40 Simulated Annealing:1000 CMA-ES:618	NelderMead : $1.1*10^{-4}$ Simulated Annealing: $5.4*10^{-9}$ CMA-ES: $1.3*10^{-12}$

Task 2 (Analysis of hyperparameter optimization on MNIST dataset by the methods):

	Function evaluations	Stability	Final accuracy
Nelder Mead	59	Good	0.15
Simulated Annealing	1000	Very Stable	0.8
CMA-ES	35	Not good	0.8

Task 3 (Final Stability Analysis):

CMA-ES is very much sensitive to initial condition given rather than the other two.

Trade-offs:

1) Nelder mead method is good for fast computation but it gives larger errors.
2) Simulated annealing is very stable and robust to initial conditions but takes too long to compute
3) CMA-ES method is a balance between both, with a bit longer time required than NM, but a better accuracy sometimes even better than simulated annealing. But it lacks stability on multiple runs and robustness. Overall, it is the best method that can be used.

Data Visualisation:

In clustering quality, diffusion maps have the best score with its silhouette being almost 0.98. But since the eigenvalues of diffusion maps are very close to each other, the top 2 or 3 eigenvalues do not reflect the nature of the data. Thus choosing top 2 eigenvectors will not give the desired result. This its ARI score is very low, 0.0158, which is still good but not better than the pca and tsne which have been directly called. The method by which diffusion maps is made makes this contradiction. Also, if we see the bincount distribution of labels given by the code, we see that there is a very uneven distribution, as a class gets almost 6000 points while the rest get 30,15 etc. Hence the problem lies in the formulation and also in the data taken. Only train data in the x direction is taken making the data a bit biased. If some other data had been used, which would have coherently described the data, it might have been better. Thus, this is the whole interpretation of the results. Another possible reason for this mismatch may be overfitting in the local space which diffusion maps are very good at.

DRW similarity matrix has been created and then invoked later in PART B. Time series is not overlapped again because the dataset is already overlapped.

Reasons for diffusion maps working well(high silhouette score):

1) It involves dtw distance which is ideal for time series data where the data might not always be well conditioned into equal lengths.
2) Reduces noise in the data because as the name suggests, it *diffuses* the incoming noise errors. Hence, time series data which is susceptible to noise, have better clustering in diffusion maps
3) Due to eigenvalue decomposition, its low dimension embedding usually catches the behavior of the data(somehow happening in my case), and hence help to classify the data better.