EE535P-SYSTEMS DESIGN

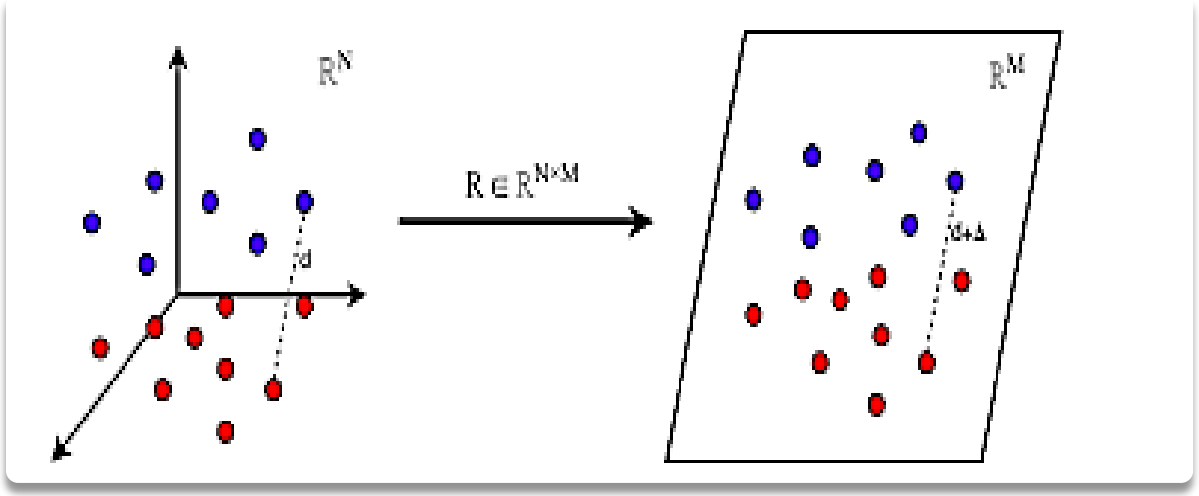
Empirical Study Of Random Projection

PRESENTED BY
Prajyot Morey T21009
M.Tech (CSP)

INSTRUCTOR
Dr. Rameshwar Pratap

CONTENTS

- Objective
- Motivation
- Expected Output
- Results and Plots
- TimeLine
- Gratitude



Picture Courtesy Internet

Objective

- To Show **Superiority of Random Projection** over Principle Component Analysis when applied Over a **mixture of gaussian in High Dimensions.**
- To Show **Projected dimension is independent** of the original dimension.
- To show that after projecting high dimensional data to low dimension the **Eccentricity of the Gaussian will decrease** and gaussian become more **spherical**.
- To find the **separation between clusters of gaussian** before and after transforming to low dimensions.

Motivation

Why Random Projection?

- PCA fails to reduce the dimensionality of mixture of 'K' Gaussian below $\boldsymbol{\Omega(K)}$ but The Random Projection will significantly reduce the dimension of the mixture of K gaussian To just $\boldsymbol{0(logk)}$ This will lead to a **significant improvement in the computational cost of Models**.

- **Time complexity** of PCA in ***'n'*** dimensional data is $\boldsymbol{0(n^3)}$ where as in Random Projection is $\boldsymbol{0(dn)}$ (where ***'d'*** is projected dimension and ***'n'*** is original dimension and ***'d'<<'n'***). Time complexity improved in Random projection as compare to PCA.

- PCA **fails to reduce the Eccentricity** of high dimensional Gaussian, this high eccentric gaussian will create problem while learning models and it become difficult to design algorithms. But **Random Projection overcome** this problems.

- Due to all mentioned Drawbacks we are **going to study behavior and result of Random Projection and will compare with PCA result**.

Expected Outputs

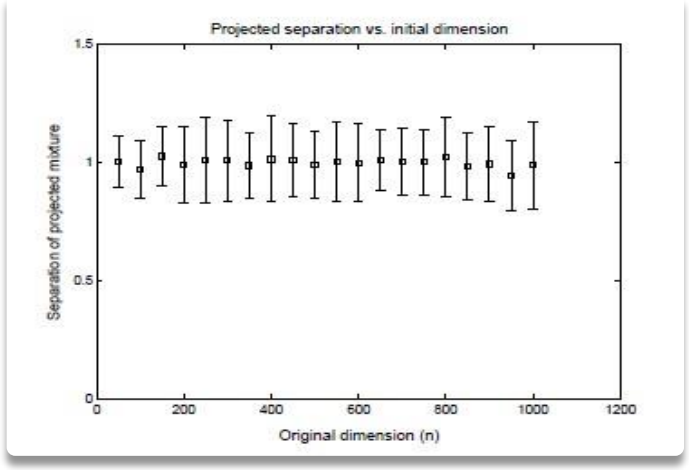


Fig. 1. The above Graphs show that the Projected dimension is independent of the original Dimension. From: Dasgupta, S. (2013), Figure 3,Experiments with Random Projection.

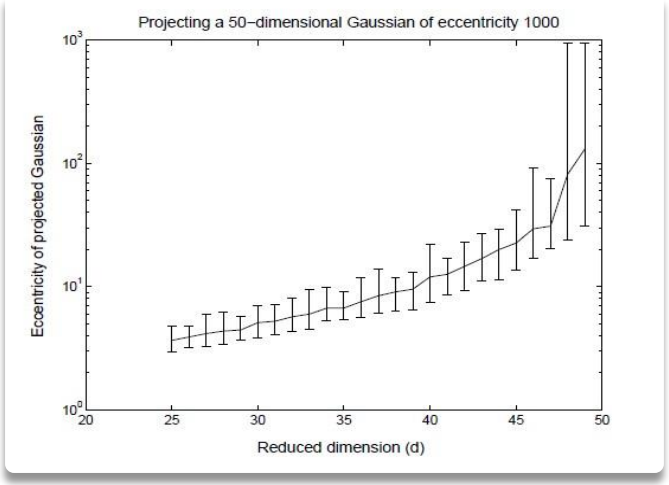


Fig.2 Above graphs show that the Eccentricity of gaussian is reduced as it is projected into low dimensions. From: From: Dasgupta, S. (2013), Figure 6, Experiments with Random Projection.

	1	2	3	4	5
1	0	0.04	0.03	0.03	0.02
2	0.04	0	0.03	0.04	0.03
3	0.03	0.03	0	0.03	0.03
4	0.03	0.04	0.03	0	0.03
5	0.02	0.03	0.03	0.03	0

	1	2	3	4	5
1	0	0.57	0.60	0.44	0.41
2	0.57	0	0.68	0.43	0.43
3	0.60	0.68	0	0.64	0.45
4	0.44	0.43	0.64	0	0.37
5	0.41	0.43	0.45	0.37	0

Fig. 3. Above figure gives idea How inter-cluster separation is affected after projecting to low dimension, left matrix represented output from PCA and right Matrix represented output from Random Projection. From: Dasgupta, S. (2013), Figure 7, Experiments with Random Projection

Results and Outputs

Experiment_1_A

Distance Between Samples before Random Projection

```
******************Distance Between Samples Before Random Projection******************

[[ 0.          9.8881654   9.99032785 ...  9.99795149  9.95279542
  10.08982287]
 [ 9.8881654   0.          9.95596273 ...  9.95673555  9.74722412
   9.78551536]
 [ 9.99032785  9.95596273  0.         ... 10.06331721  9.89426062
  10.03954993]
 ...
 [ 9.99795149  9.95673555 10.06331721 ...  0.         10.01121362
   9.89763293]
 [ 9.95279542  9.74722412  9.89426062 ... 10.01121362  0.
   9.91200757]
 [10.08982287  9.78551536 10.03954993 ...  9.89763293  9.91200757
   0.        ]]

 Dimension of Distance Matrix is :- (10000, 10000)
```

Distance Between Samples After Random Projection

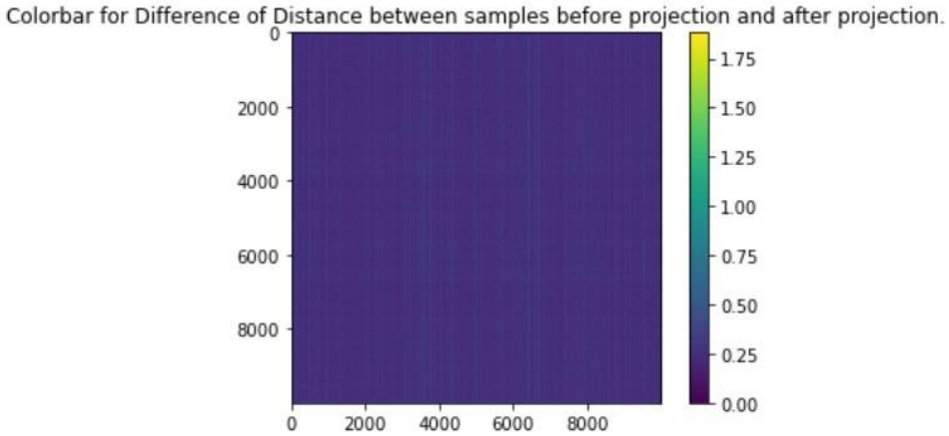
```
**********Distance Between Samples After Random Projection**********

[[ 0.          9.42512614  9.60454831 ...  9.43715107  9.50393203
   9.96562294]
 [ 9.42512614  0.          9.85112024 ...  9.59942006  9.24011401
   9.33766046]
 [ 9.60454831  9.85112024  0.         ...  9.69377678  9.87891802
  10.77679372]
 ...
 [ 9.43715107  9.59942006  9.69377678 ...  0.          9.32574203
   9.67893823]
 [ 9.50393203  9.24011401  9.87891802 ...  9.32574203  0.
  10.0586824 ]
 [ 9.96562294  9.33766046 10.77679372 ...  9.67893823 10.0586824
   0.        ]]
```

Difference between the distance of samples before and after projection.

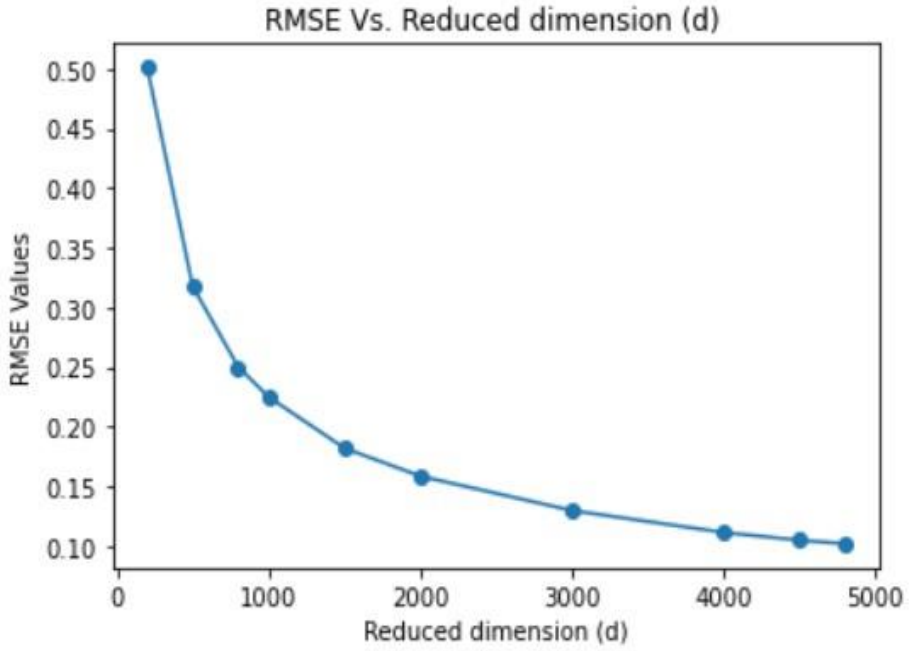
Difference Between distance of samples before projection and after projection.

```
[[0.         0.46303926 0.38577955 ... 0.56080042 0.44886339 0.12419993]
 [0.46303926 0.         0.10484249 ... 0.35731549 0.50711012 0.44785491]
 [0.38577955 0.10484249 0.         ... 0.36954043 0.0153426  0.73724379]
 ...
 [0.56080042 0.35731549 0.36954043 ... 0.         0.68547159 0.2186947 ]
 [0.44886339 0.50711012 0.0153426  ... 0.68547159 0.         0.14667483]
 [0.12419993 0.44785491 0.73724379 ... 0.2186947  0.14667483 0.        ]]
```

Heatmap for difference matrix.



Experimen_1_B&C



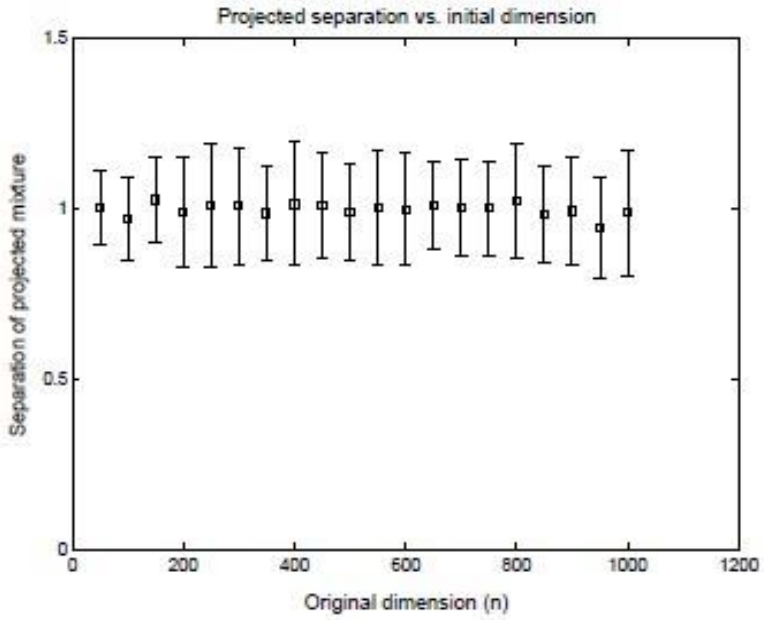


Fig. 4. The above Graphs show that the Projected dimension is independent of the original Dimension. From: Dasgupta, S. (2013), Figure 3, Experiments with Random Projection.

Experiment_2

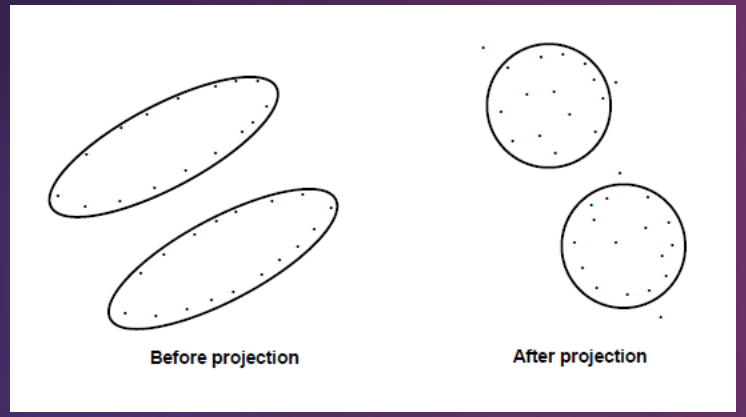
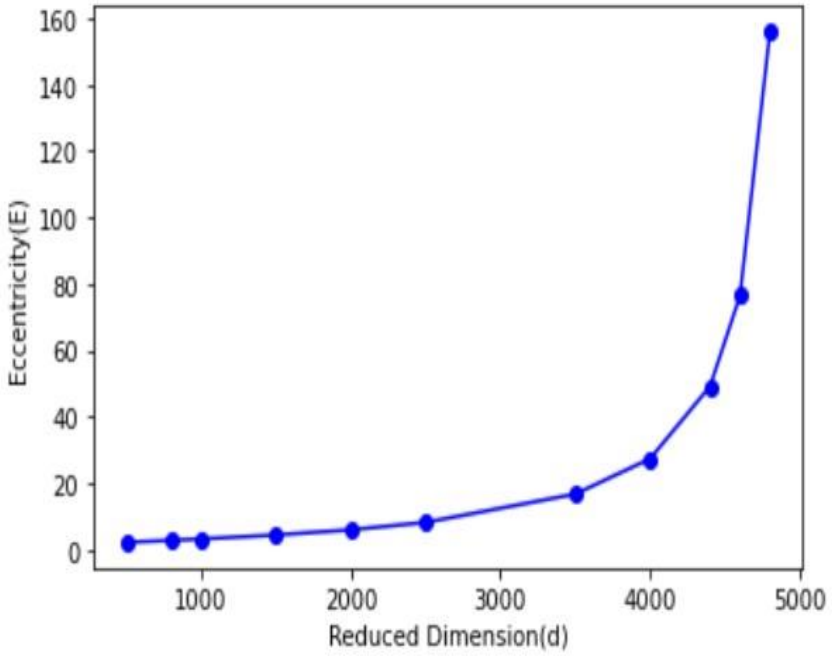
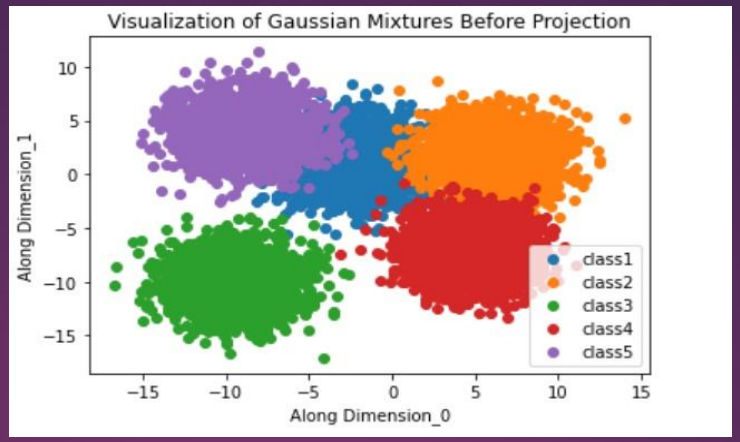
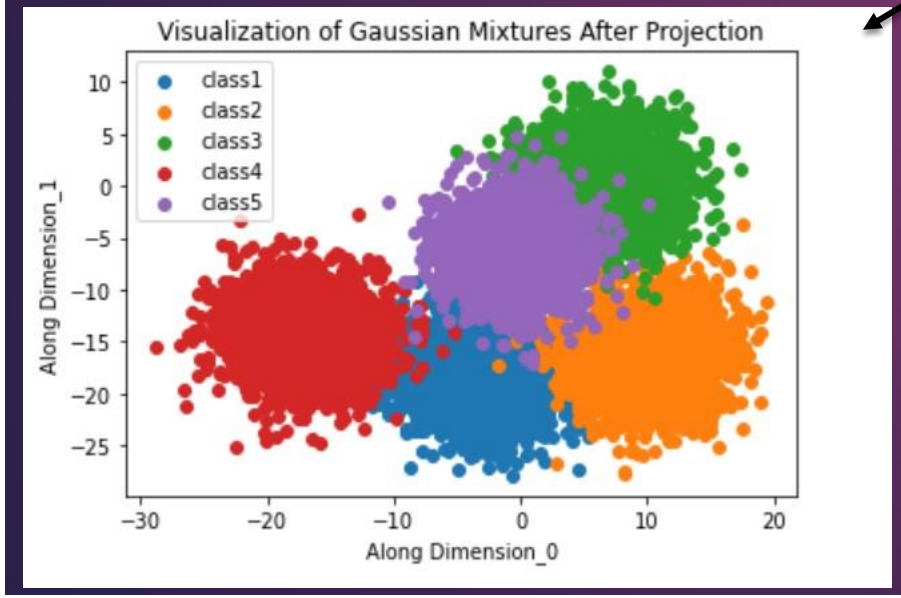


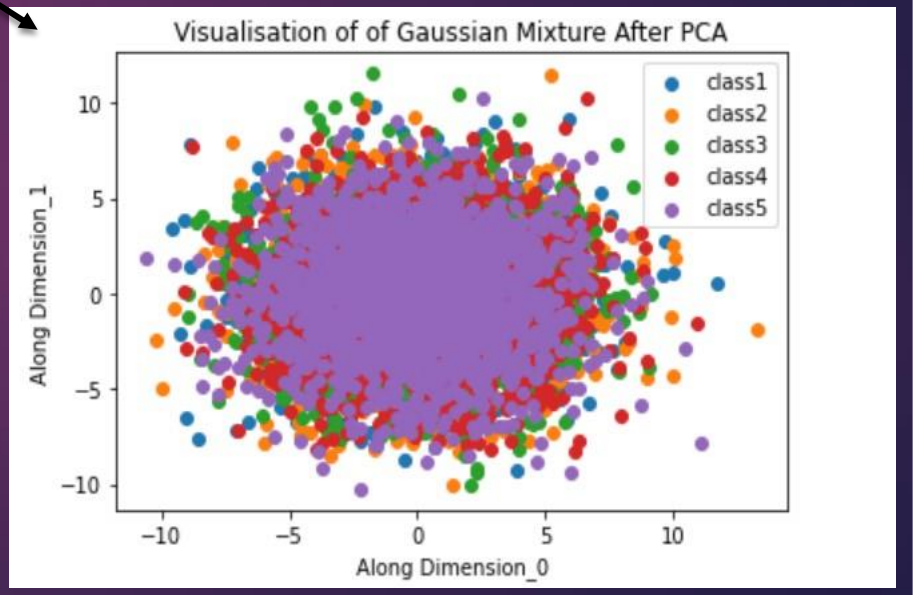
Figure 5: The effects of random projection: the dimension is drastically reduced while the clusters remain well-separated become more spherical. From: Dasgupta, S. (2013), Figure 2, Experiments with Random Projection.



Experiment_3



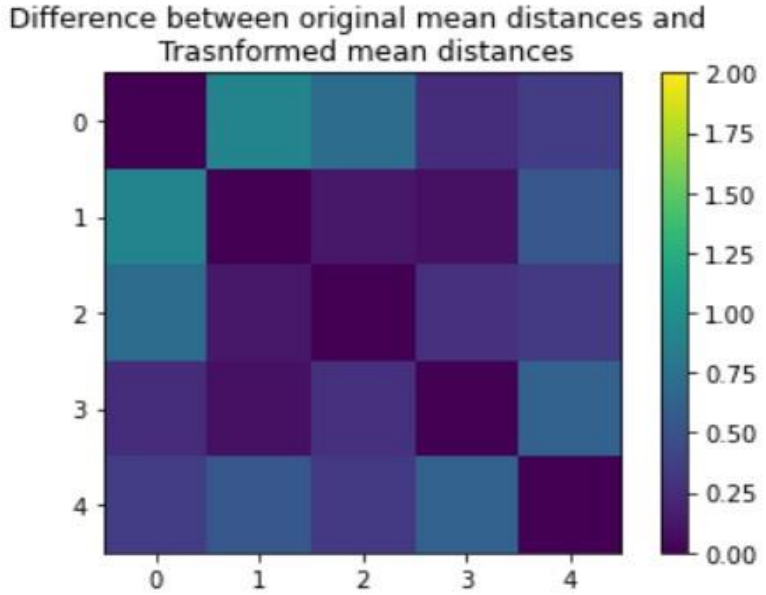




After Applying Random Projection.

Difference between means of Gaussian pair before and after projection.

```
[[0.         0.90527967 0.70378808 0.25966286 0.3660051 ]
 [0.90527967 0.         0.128858   0.08602527 0.53927458]
 [0.70378808 0.128858   0.         0.28751687 0.33778298]
 [0.25966286 0.08602527 0.28751687 0.         0.62529985]
 [0.3660051  0.53927458 0.33778298 0.62529985 0.        ]]
```



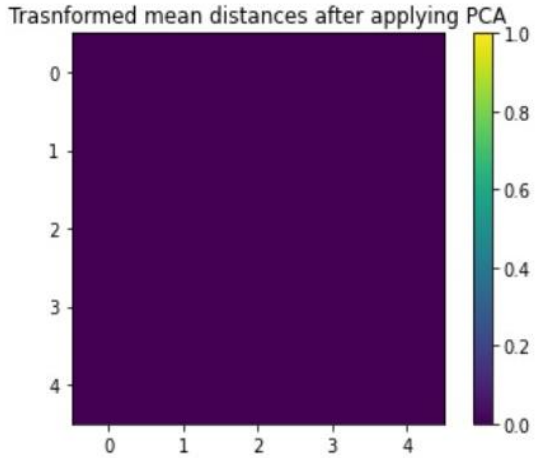
After Applying PCA.

Distances between the mean of Gaussian mixture before PCA.

```
[[0.         0.59112373 0.62744052 0.36582104 0.27719344]
 [0.59112373 0.         0.0363168  0.95694477 0.31393029]
 [0.62744052 0.0363168  0.         0.99326157 0.35024709]
 [0.36582104 0.95694477 0.99326157 0.         0.64301448]
 [0.27719344 0.31393029 0.35024709 0.64301448 0.        ]]
```

Distances between the mean of gaussian mixtures after applying PCA.

```
[[0.00000000e+00 1.22888366e-16 1.10091491e-15 5.39444045e-16
  8.04014633e-16]
 [1.22888366e-16 0.00000000e+00 1.22380328e-15 6.62332411e-16
  9.26902999e-16]
 [1.10091491e-15 1.22380328e-15 0.00000000e+00 5.61470870e-16
  2.96900282e-16]
 [5.39444045e-16 6.62332411e-16 5.61470870e-16 0.00000000e+00
  2.64570588e-16]
 [8.04014633e-16 9.26902999e-16 2.96900282e-16 2.64570588e-16
  0.00000000e+00]]
```



TimeLine

17 –Dec To 31-Dec

- Introduction
- Prerequisite Knowledge
- Started Reading about JL
- Understand JL lemma proof.
- Understand Working of random Projection

1- Jan To 15-Jan

- Presentation to the instructor about an understanding of JL lemma and Random projection.
- Confirmation from instructor
- Project Problem Statement assigned.
- Started reading and researching about Problem statement.

16-Jan To 31-Jan

- Mid Evaluation Documentation work.
- Working on the algorithm and experiment part.
- Generation of synthetic dataset and experimenting.

1-Feb To 15-Feb

- Worked on the Algorithm part.
- Did series of Experiments as per objective.
- Documentation Part for Final Evaluation
- Create a GitHub repository for project work.

Thank You…!