EE535P Systems Design

Empirical Study Of Random Projection

Prajyot Morey T21009 M.Tech (CSP)

Instructor:
Dr. Rameshwar Pratap

19-Jan-2022

INDIAN INSTITUTE OF TECHNOLOGY MANDI
Mandi-175001, India
School Of Computing and Electrical Engineering

February, 2022

Contents

List of Figures

1 Objective

The objective of our project is to show the superiority of the Random Projection algorithm in comparison to PCA algorithms when applied to a High dimensional Gaussian mixture data to transform into low-dimensional Gaussian mixture data. To conclude the results and observation series experiment will be carried out on synthetic data first, then after learning results and observation from experiments on synthetic data another experiment will be carried out on the real-world data to comment on the accuracy of models. experiment which we are going to perform are as follows,

• Experiment 1_A, This Experiment is performed to show the working of JL-Lemma theorem, which is the main working principle behind the Random Projection.

• Experiment 1_B, This experiment is intended to show that projected dimension **d** does not depend on the original dimension **n**. And to find out the optimum range for low dimensions projection for Random projection.

• Experiment 2, to comment on Eccentricity of the Gaussian mixtures cluster before and after a projecting to low dimension in Random Projection.

• Experiment 3, to comment on the intercluster separation between a group of clusters before and after projection to low dimension in both PCA and Random Projection.

After performing all the above experiments. One can deduce and learn the behavior and result of PCA and Random Projection and comment on the Superiority among both algorithms. after studying and observing the results another experiment will be performed in which we are planning to design a classification model on real-world data using both PCA and Random Projection as dimensionality reduction algorithms and observe and deduce the accuracy of both models. and comment on the superiority of both models.

2 Introduction

In today's internet world, an abundant amount of data is available on the internet due to technologies and platforms like Social Media, IoT, Online transactions Data, Cloud services, and High definition cameras, and many more. These data are rich in information and high quality, which results in the dimensionality of this data being very high, by using these data in technology like Artificial intelligence and applying machine learning algorithms on these high dimensional data for learning purposes is time-consuming and memory inefficient. also, it faces problems like "Overfitting" and "Curse of dimensionality", to overcome these problems Dimensionality reduction methods are used, Dimension reduction methods compress a large space of high dimensional data onto a new subspace of lower-dimensional without losing the important information. For example:- If a single point in a dataset is of size 'd' then it can be compressed to the size of 'k', where d can be in billions and k can be of size in thousand, precisely k « d. this will result in significant improvements in time consumption and memory efficiency and become very cost-effective, there are many dimensionality reduction algorithms present like PCA, SVD, LDA, and Random Projection. All these dimensionality reduction algorithms preserve similarity measures in a dataset after compressing to lower dimensions these similarity measures can be angular similarity, Euclidean distance, or variance and covariance.

Random Projection is a promising dimensionality reduction technique for learning a mixture of gaussian. Random projection transform high-dimensional data to lower-dimensional data by preserving the distance between any two points. The seminal work of [Johnson and Lindenstrauss, 1983] [1] also known as the JL lemma, compresses real-value high dimensional datasets to lower-dimensional real-value datasets and also preserves the pairwise euclidean distance between any pair of points in a low dimensional dataset.

Recent theoretical work by (Dasgupta, 1999) [2]suggests that for learning of high dimensional mixture of gaussian we can transform high dimensional Gaussian mixture data to low dimensional subspace by using Random projection as dimensionality reduction. this will lead to gives good results. there are two theoretical results about random projection after transforming to low dimensional subspace, our Aim is to do an empirical research study on these results and design a classification model to demonstrate results. the first result suggests that data from a Mixture of K Gaussian can be projected into a subspace of low dimension, just O(log k)[3]. And the second one is, after projecting to the low dimension subspace the distance between gaussian clusters will be the same before and after projection[3]. This projected dimension is independent of the number of the original dimension. Before projecting data, Due to the high dimensionality, clusters of gaussian are highly eccentric, this high eccentricity of gaussian becomes an algorithmic challenge i.e., EM algorithms face problems while learning gaussian, and also classification models get confused in high eccentric gaussian data[3]. The second results suggest that after transforming high dimensional data to a low dimensional subspace, the eccentric gaussian will become more spherical[3].

These two benefits of random projection help to achieve improvement in time complexity while learning the Gaussian model and increase the accuracy in classification, for demonstrating this result and the benefits of random projection algorithm, we are going design a classification model for OCR from a canonical USPS dataset using random projection as dimensionality algorithm and compare this result with famous dimensionality algorithm PCA.

2.1 Motivation

Principle component analysis is one of the famous dimensionality reduction algorithms, one of the goals of our project is to do a comparative study of PCA and Random Projection. we try to overcome the abnormalities face by the PCA while applying on the mixture of Gaussian. Principal component analysis (PCA) is used to reduce the dimensionality of data by retaining vector components that have the maximum variation along its direction and discarding an insignificant vector component i.e., vectors that have the least variation along a vector. These maximum variation components are nothing but the principle components. the theoretical study suggests that PCA can not reduce the dimension of data from the mixture of K gaussian to below $\Omega(K)$ dimension, whereas random projection can perform better than this result and reduce the dimension to just O(log K). this leads to high time complexity in PCA $O(n^3)$ where n is the dimension of the data points) as compared to random projection. after transforming Gaussian mixture to reduced dimension using PCA. the theoretical study suggests that PCA will collapse all clusters into one and didn't maintain the intercluster separation, due to which classification and learning model gets confused which lead to an increase in the inaccuracy of the ML model and this will be a serious issue in developing models, and other problems face by PCA is, as of today's internet world data present in very high dimension due to plenty of sensors and sources, this High dimensional data cluster become eccentric this eccentricity become an algorithmic challenge, i.e., EM algorithms face problems while learning, also classification models will get confused in high eccentricity. It is conceptually much easier to design an algorithm for a spherical cluster than ellipsoidal, PCA measurably fails to reduce the eccentricity of transformed data. one of the reasons behind this is, PCA gets fooled into picking directions that correspond to the eccentricity of individual gaussian and capture very little of the intercluster variation.

The seminal work from [Sanjay Dasgupta, 2013] [3] suggests that Random Projection will overcome all the problems stated about PCA when applied to the mixture of gaussian. The experimental study suggests that Random Projection will project a mixture of K gaussian to just O(log K), which is relatively very low than PCA, and the time complexity will be O(d*n) where d is reduced and n is the original dimension. Which is much lower as compared to PCA. The projected gaussian mixture also resulted in low eccentricity and a well-separated cluster of gaussian this will become easier to design algorithms for spherical and well-separated clusters. all these results and benefits of random projection will give better results and reduce the pain to design learning and classification algorithms.

3 Problem and System Description

3.1 Problem

To carry out a series of experiments to study the empirical properties and results of Random Projection. after studying and observing results from experiments we have commented on the uses and advantages of Random Projection as a dimensionality algorithm as compared to other Dimensionality algorithms.

3.2 System Description

Experimental Setup

To carry out experiments of our objectives, we need Synthetic data that we will make using python's sklearn library. and for real-world data for designing OCR classifier, we are going to use the USPS dataset of Handwritten digits. and will perform a series of experiments on synthetic data followed by designing a classifier using a real-world USPS dataset. to conclude and comment on observation and results.

Synthetic data of various original dimensions are generated. like, n=100 i.e. R^{100}, n=200 i.e. R^{200}, n=400 i.e. R^{400}. These synthetic data are being projected into various low dimensions according to the need of our experiment (Different projected dimensions for different Experiments) e.g., 40,25,20.... and so on. Real Word data from the USPS dataset contain handwritten digits. The training data consist of 9,709 labeled instances of handwritten digits. each digit is represented as a vector in $[-1,1]^{256}$ and each image is a 16*16 bitmap image. These data are being used in the experiment.

4 Theorem, Proof and Lemma Used In Project

1. JL lemma. [Johnson and Lindenstrauss, 1983][1]
Random Projection is a dimensionality reduction algorithm, which uses the JL lemma,[Johnson and Lindenstrauss, 1983] [1] as the main working principle. This lemma's result is used in the first Experiment of this project where we have to show that separation between any two points is near about constant after projection.

Theorem 1 *Given $0 < \varepsilon < 1$, a set X of m in $\mathbb{R}^n$ and, a number $n > 8\ \frac{ln(m)}{\varepsilon^2}$, there is a linear map $f : \mathbb{R}^N \rightarrow \mathbb{R}^n$ such that,*

$$(1-\varepsilon)\|u-v\|^2 \leq \|f(u)-f(v)\|^2 \leq (1+\varepsilon)\|u-v\|^2$$

for all $u, v \in X$

2. Lemma for separation between the mean of gaussian after projection to low dimension.

The following dimensionality reduction lemma applies to arbitrary mixtures of Gaussians. And implies that by random projection, data from a mixture of 'K' Gaussians can be mapped into a subspace of dimension just O(log K) [2]. This lemma has been used in experiment No.3 of this project. Experiment No. 3 is carried out to study the difference between the mean separation of gaussian before and after projection.

Lemma 1 (Dasgupta, 1999) *[2] For any $c > 0$, pick a c-separated mixture of k Gaussians in R^n . Let $\delta, \epsilon \in (0, 1)$ designate confidence and accuracy parameters, respectively. Suppose the mixture is projected into a randomly chosen subspace of dimension $d \geq \frac{C1}{\xi} ln\frac{k}{\delta}$, where C1 is some universal constant. Then, with probability $> 1-\delta$ over the choice of subspace, the projected mixture in R^d will be $(C1\sqrt{1-\epsilon})$-separated*

3.Definition of Eccentricity.

After applying Random Projection we show that even if the original Gaussians are highly skewed (have ellipsoidal contours of high eccentricity), their projected counterparts will be more spherical. Since it is conceptually much easier to design algorithms for spherical clusters than ellipsoidal ones, this feature of random projection is very useful for learning the projected mixture. This result is used in Experiment No. 2 of our project In which the following eccentricity definition is used as a reference for calculating the Eccentricity of the dataset.

Definition 1 *For a positive definite matrix Σ,let $\lambda\_max(\Sigma)$ and $\lambda\_min(\Sigma)$ refer to its largest and smallest eigenvalues, respectively, and denote by $E(\Sigma)$ the eccentricity of the matrix, that is, $\sqrt{\frac{\lambda\_max(\Sigma)}{\lambda\_min(\Sigma)}}$*

5 Experiment Descriptions And Results

5.1 Experiment 1

Experiment_1_A:- JL-Lemma

Experiment 1_A is performed to show the working of the JL lemma. Here we have taken a dataset of 10000 samples and 5000 features with a standard deviation of 0.1. This is synthetic dataset which we have generated by using the make_blobs() function of sklearn library. As per the JL lemma, the euclidean distance between any two points before and after projection is constant with some error range, i.e esp= 0.5. Euclidean distance between samples is calculated before and after projection, and results are plotted and observed.

Results and Plots

.

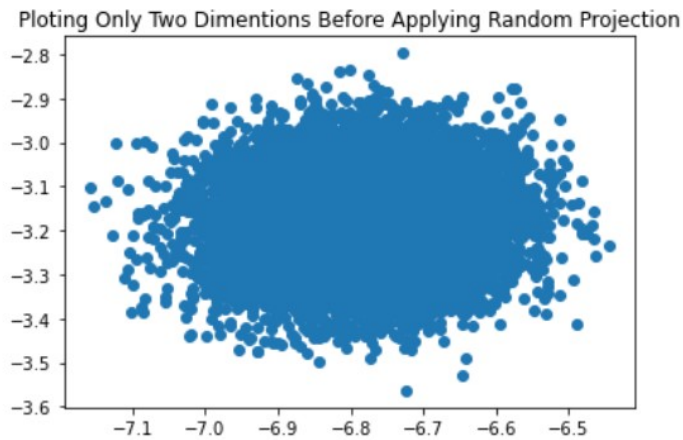


Figure 1: Data Distribuation Without Reducing To Low Dimension.

```
*******************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)
```

Figure 2: Distances Between sample point of original dimension before applying Random Projecction.

```
**********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.        ]]
```

Figure 3: Distances Between sample point of Reduced Dimension

```
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.        ]]
```

Figure 4: Difference Of Distance between samples before and after Projection

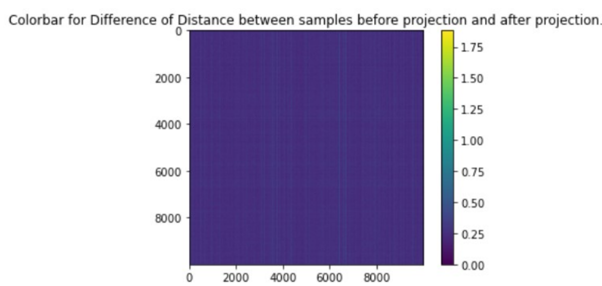


Figure 5: Difference Of Distance between samples before and after Projection

Experiment_1_B:- Experiment to find the relationship between RMSE values and Reduced Dimension.

This experiment is carried out to find the relation between RMSE(Root Mean Squared Value) of the error of the euclidean distance matrices of the sample before and after projection. For performing the experiment dataset of 10000 samples with 5000 samples is used. Random Projection is applied to the given dataset to reduce its dimension. the experiment is carried out for various values of reduced dimension and a Plot is generated for a range of reduced dimension values. After the result by observing, we can conclude that, as the reduced dimensions increase the RMSE value is decreased.

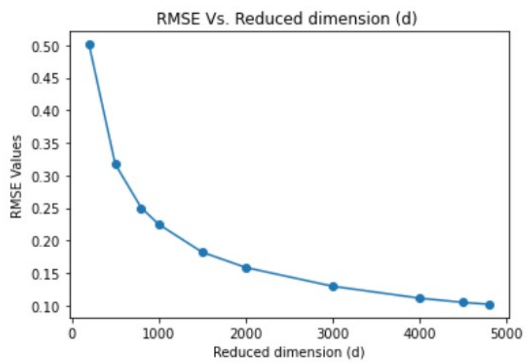


Figure 6: Relation between Reduced dimension (d) and RMSE of error between distance of samples before and after projection.

Experiment_1_C

This experiment is intended to show that projected dimension **d** does not depend on the original dimensions **n**. Random projection from **n** dimension to **d** dimension is represented by a **d*n** matrix. it does not depend on the original dimension. each entry of the matrix is an i.i.d. N(0,1).

$$\begin{bmatrix} R \end{bmatrix}_{(k\times d)} \times \begin{bmatrix} Q \end{bmatrix}_{(d\times n)} = \begin{bmatrix} Q' = R\times Q \end{bmatrix}_{(k\times n)}$$

Here two 1- separated spherical Gaussian are projected into R^{20} and their separation is noted as a function of n[3]. We are going to perform a minimum of 20 trials for each value of n. the below graph between the projected separation vs. initial dimension shows the result and we can conclude that Projected dimension d does not depend upon the original dimension[3].

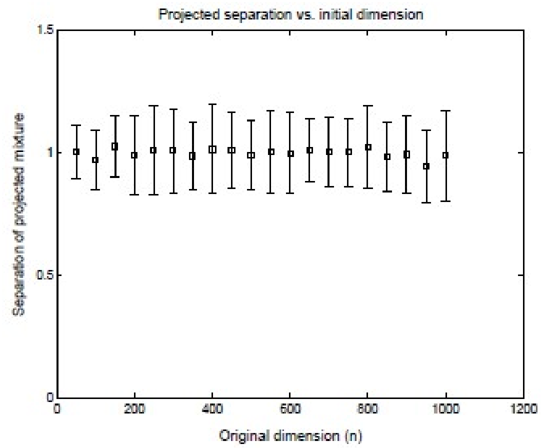


Figure 7: The projected dimension (d = 20) does not depend upon the original dimension.

5.2 Experiment 2

Experiment to find out the relationship between eccentricity and reduced dimension.

This experiment is carried out to find the relationship between eccentricity and reduced dimension. This experiment shows the decrease in the eccentricity after projecting to a low dimension using Random Projection. In this experiment we project a Gaussian of high eccentricity E from $R^{n=5000}$ into R^{500} and measure the eccentricity E^* of the projected Gaussian. In this experiment, data is projected into successively lower dimensions of 500, 800, 1000, 1500, 2000, 2500, 3500, 4000, 4400, 4600, 4800, etc.

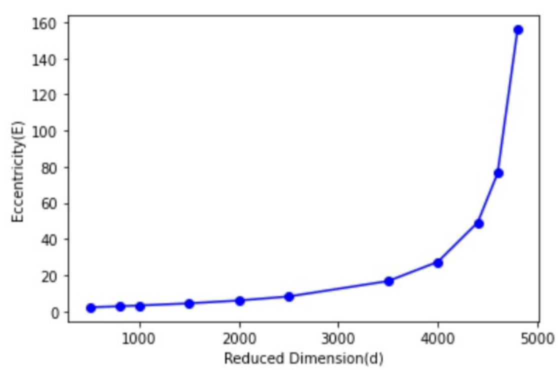


Figure 8: Eccentricity of Projected Gaussian Vs. Reduced Dimension(d)

5.3 Experiment 3

Experiment to calculate the separation between the mixture of Gaussian clusters by using Random Projection and PCA.
The seminal work from [Sanjay Dasgupta, 2013] [3] suggest that after applying Random Projection to the mixture of gaussian, the separation between the mean of gaussian will remain constant with some error. but this is not happening with PCA when we use it as a dimensionality algorithm. To stufy this result we have performed the experiment and observed the result. So for that purpose, Here a Dataset of 5 different gaussian with 10000 samples and 500 features is taken. The Euclidean distance between the mean of each gaussian is calculated before projection and after projecting data to a low dimension. The data is projected to low dimensions by using Random Projection and PCA as dimensionality reduction algorithms. After observing the result we can conclude that, **1.** After projecting data to low dimensions by using Random Projection the euclidean distance between the mean of the Gaussian cluster will remain almost the same. and **2.** After projecting data to low dimensions by using PCA the euclidean distance between the mean of gaussian will reduce drastically i.e., all the gaussian clusters will collapse into one and the mean will coincide with each other.

Results and Plots

1. For Random Projection.

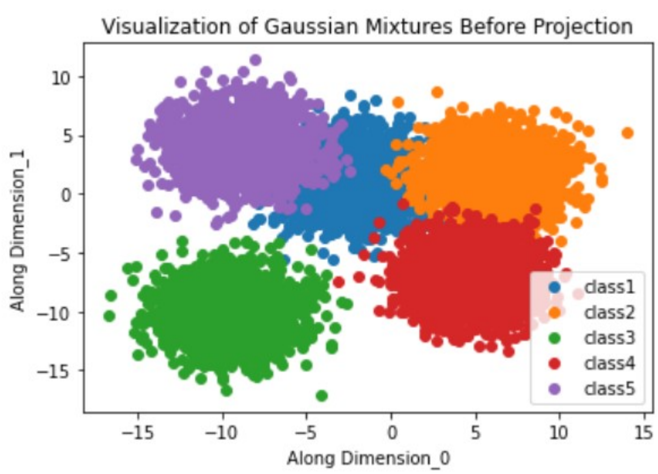


Figure 9: Mixture of Gaussian in High dimension before projecting to Low dimension.

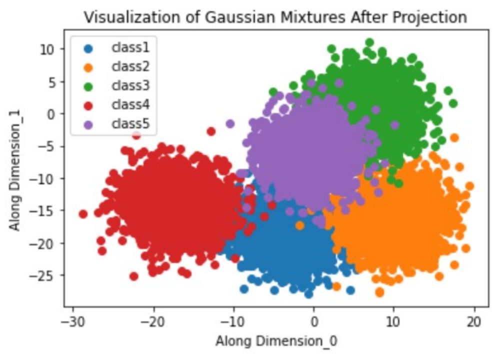


Figure 10: Mixture of Gaussian after projecting to Low dimension By using Random Projection.

```
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.        ]]
```

Figure 11: Difference between Projected Gaussian Mixtures Pair Mean and Original Gaussian Mixtures Pair Mean When Random Projection is Used.

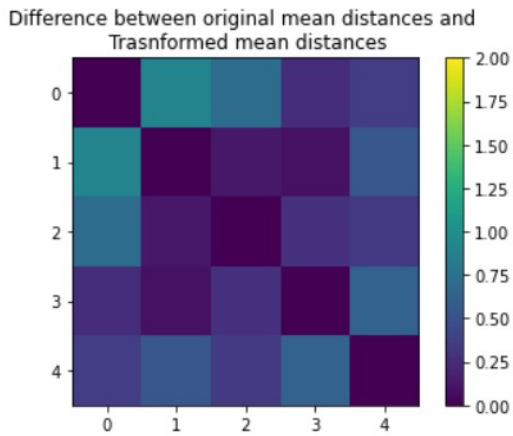


Figure 12: Difference between Projected Gaussian Mixtures Pair Mean and Original Gaussian Mixtures Pair Mean When Random Projection is Used.

2. For PCA.

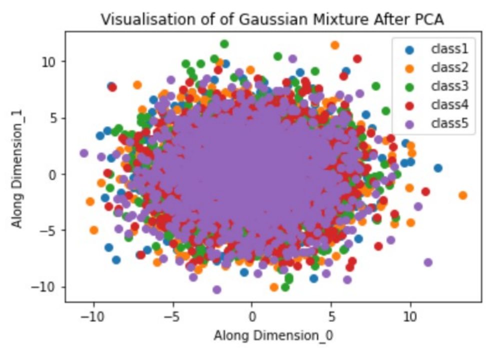


Figure 13: Mixture of Gaussian after projecting to Low dimension By using PCA.

Below Colour bar and resultant data matrix suggests that, after applying PCA on a mixture of gaussian as a dimensionality reduction algorithm, the distance between the mean of the gaussian mixture will reduce drastically to zero and all the gaussian will collapse into one.

```
[[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]]
```

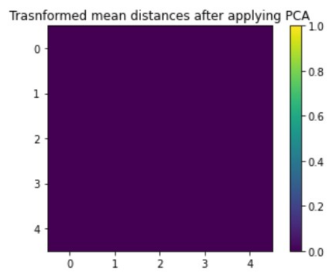


Figure 14: Colorbar and Mean Distance Matrix after Projecting data to low dimension using PCA.

6 Conclusion

Dimensionality reduction or feature extraction is one of the important steps while designing ML, DL models, it leads to improves time as well space limitations. In this project, we have concluded that Random Projection is one of the best dimensionality algorithms for high dimensional and large datasets. it overcome all the problems faced by PCA when applied to a large dataset.

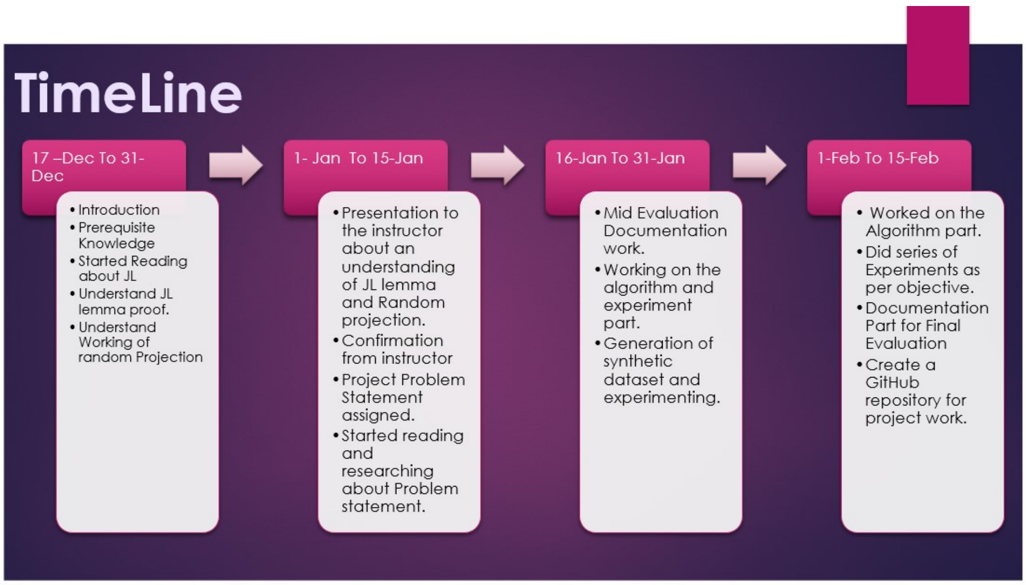
From this project, we can conclude that the reduced dimension is independent of the original dimension and only depends on the logarithm of the original samples in the dataset. also, the ambiguities faced by the PCA algorithms are overcome by Random Projection. After using random projection the reduced dataset becomes more spherical which leads to bits of help in designing the learning algorithm. one of the advantages of Random Projection is that while training the mixture of gaussian if we apply PCA as dimensionality algorithm then all the gaussian clusters will collapse into one cluster and result in the model will get confused while training, the random projection will overcome this problem and maintain the intercluster near about constant.

Till now we have seen the advantages of Random Projection, But Random Projection works as per our expectation when only we have large datasets(High dimension and High samples). Random Projection failed miserably when applied on low dimension small datasets. as we have experimented all the experiments on Gaussian distributed mixture model datasets, but nowadays many ML problems used other types of distribution mixture models, eg. Tree-structured distribution. researched are going on to study the properties and result of Random Projection when applied to other distributed models.

7 Project Github Page

https://github.com/prajyotmorey/EE535P-Systems-Design_Random-Projection

8 TimeLine



Bibliography

[1] P. Frankl and H. Maehara, "The johnson-lindenstrauss lemma and the sphericity of some graphs," *Journal of Combinatorial Theory, Series B*, vol. 44, no. 3, pp. 355–362, 1988, ISSN: 0095-8956. DOI: https://doi.org/10.1016/0095-8956(88)90043-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0095895688900433.

[2] S. Dasgupta, "Learning mixtures of gaussians," in *40th Annual Symposium on Foundations of Computer Science (Cat. No.99CB37039)*, 1999, pp. 634–644. DOI: 10.1109/SFFCS.1999.814639.

[3] S. Dasgupta, *Experiments with random projection*, 2013. arXiv: 1301.3849 [cs.LG].