CCE Proficience 2022

Project presentation ,Computing for AI ML

Nov 29 , 2022

By Shristi Singh, Parinita Bora,

Introduction

This addresses

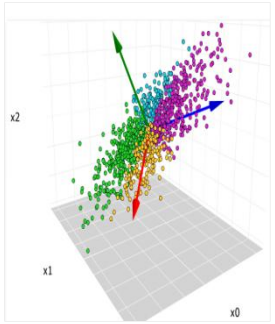
- **Introduction to the project**
  Topic : Image Compression using PCA, SVD and K-Mean algorithm

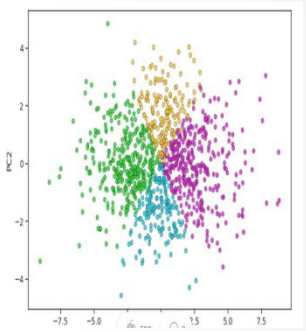
- **Apply the algorithms on the data as a part of preprocessing task and experiment for data compression**

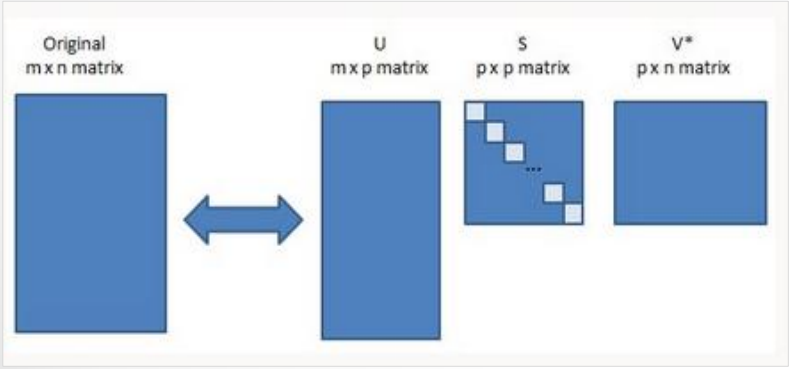
Data of high dimension for machine learning

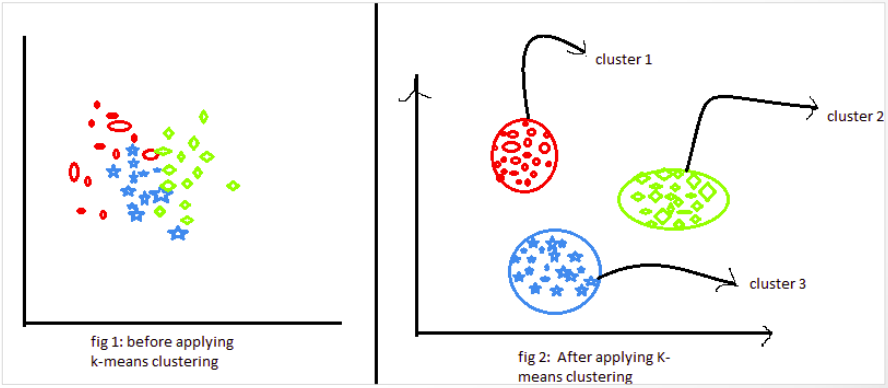
Unsupervised algorithms & preprocessing

- Dimensionality reduction.
- Visualization and exploratory analysis
- Reduced complexity of Model
- Reduced training time
- SVD, PCA, K-mean









Material and methods

Data : High resolution image

fig.3 . Image of a Cosmic object, captured by James Webb Space Telescope (publicly available in NASA website) dimension 570X 985 x 3

Material and methods

PCA	SVD	K-mean
Step1. Calculate the covariance matrix of the data	Step1. getting three component matrices with Red , Blue and green constituents	Step 1. An optimal number of clusters (K) is chosen.
step2. Extract the eigenvectors and the eigenvalues of that matrix	Step2. Applying SVD on each of the three components to generate three vectors for each of the matrices	Step 2. k number of points "centroids" are initialized randomly within the data area.
Step3. Select the number of desired dimensions and filter the eigenvectors to match it, sorting them by their associated eigenvalue	Step3. Preserving only K i.e. Selecting k columns from U matrix and k rows from VT matrix, and resetting rest to zero	Step 3. Each data or observation is attributed to own closest centroid.
Step4. Multiply the original space by the feature vector generated in the previous step.	Step4. Reconstructing the coloured components from U and V	Step 4. Updating is done for the centroids to hold the value corresponding to the centre of its all attributed observations.
	Step5. Final image is formed by concatenating the three components	

Results and Discussion

#components(Principal component) /component SVD/ cluster for K-mean	Compression ratio (%) PCA	Compression ratio (%) SVD	Compression ratio (%) K-Mean
10	99.08	97.23	98.25
20	98.15	94.46	96.49
30	97.23	91.69	94.74
40	96.30	88.91	92.98
50	95.38	86.14	91.23
60	94.46	83.37	89.47
70	93.53	80.60	87.72
80	92.61	77.83	85.96
90	91.69	75.06	84.21
100	90.76	72.29	82.46

Results and Discussion

Compression ratio = ((original_number_of_image_element -new_number_of_values after applying the algorithm)/original_number_of_image element)*100.

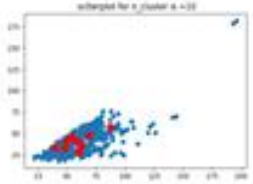
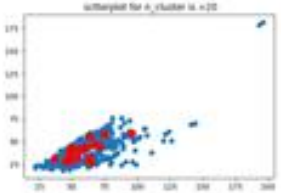
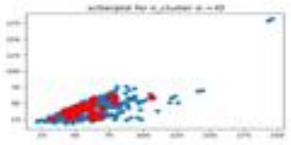
Repositories :

PCA	**https://drive.google.com/file/d/1_pBJL6v9sRRetdD0tLqvmihOVvtvivf8/view?usp=share_link**
SVD	https://colab.research.google.com/drive/1eG843MHVTwohPAqRmsQa8JToxPNJZR1M?usp=share_link
K-Mean	https://drive.google.com/file/d/1VFxHAb34riaiYDiqaN0uYt8Jw4hqJbUk/view?usp=sharing

Github: https://github.coom/Gitpabora/Data_reduction_compression

Results & Discussion

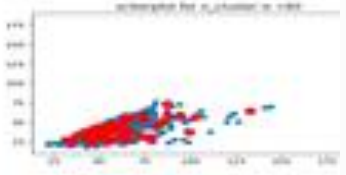
Table 5 Reconstructed Images for PCA & SVD , Kmean clustering

#components(Principal component) /component SVD/ cluster for K-mean	PCA reconstructed image	Reconstructed Image after SVD	K-mean Scatter plot
10			
20			
30			
40			

Results & Discussion

50

60



70

Results & Discussion

80

90

100

Results & Discussion

1. Note: The image for K mean clustering is placed only showing the cluster formation, not comparable in terms of reconstruction.

2. In both the algorithms for PCA and SVD as the Number of principal component or K the compression ratio decreases.

2. Reconstruction for PCA is better at a lower value of number of principal components
3. The compression ratio higher in PCA for the same value of component in PCA and K value in SVD

FUTURE SCOPE

1. Exploring other data reduction techniques for Machine learning.
2. Most importantly

(a) Experimenting with large dataset and setting up github CI
(b) test for the measures of these algorithms in terms of the impact on the model performance (c) when which algorithm is suitable.

The applicability which algorithm is most appropriate can only be experimented after evaluating accuracy of the model for the pre-treated data by these algorithms

13. References

Omar H.D et.al. (2020). Algeria Image Compression using PCA, 2021, 1–11. 2020 International Conference on Mathematics and Information Technology

Blog https://towardsdatascience.com/pca-102-should-you-use-pca-how-many-components-to-use-how-to-interpret-them-da0c8e3b11f0

https://courses.grainger.illinois.edu/cs357/fa2020/assets/lectures/complete-slides/18-PCA.pdf

https://scikit-learn.org/stable/modules/clustering.html

James Fowler, , Qian Du , IEEE member , Hyper spectral Image Compression using JPEG2000 and PCA

Stewart, W. G. (1993)On the early History of the Singular Value Decomposition, SIAM Review, Volume 35, Issue 4. Dec, 1993 551-556 , JTHOR, available at https://www.math.ucdavis.edu/~saito/courses/229A/stewart-svd.pdf

Toderici, G. (2017), Full Resolution Image Compression wit Recurrent Neural Network , https://arxiv.org/pdf/1608.05148.pdf

Vieira, M.N.C.S Vasco (2012)Permutation tests to estimate significance on Principal component analysis ., Technical university of Lisbon , available at https://www.researchgate.net/publication/255728363

14. 'Tha… (1 slide)

Jus… le.

THANK YOU

Questions?