Overview

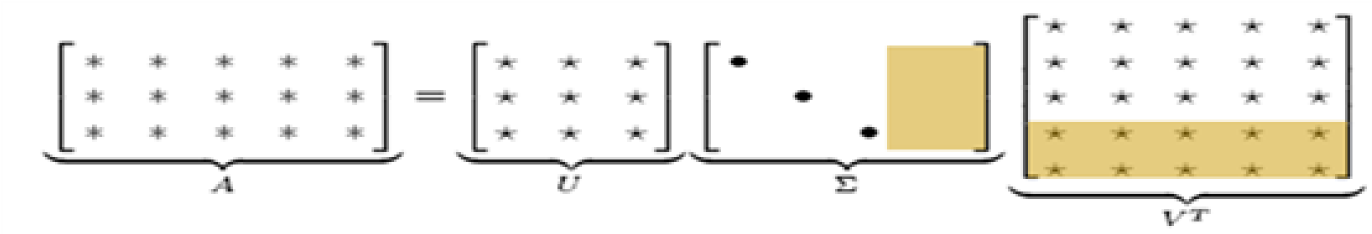
- What is SVD
- Steps done
- SVD Compressed Images
- Observations
- Demo

- PCA
- Dataset
- Data Processing
- Results of PCA

SVD

Any m by n matrix A of rank r can be factorized in the form of $A = U\Sigma V^T$ where the columns of V are orthogonal eigenvectors of A^TA and the columns of U are orthogonal eigenvectors of AA^T .

The singular matrix $\Sigma = diag(\sigma_1 ... \sigma_r)$ where $\sigma_1 ... \sigma_r$ are the singular values which are square roots of the eigen values of AA^T .



The lower rank of matrix can be used to eliminate the eigen values in S matrix and corresponding columns of U and V.

Steps Followed

- Step1: Converted the ASCII image into Binary
- Step2: Stored the U ,V and singular values in binary format after eliminating the singular values using rank approximation
- Comparing the sizes of images obtained in step1 and 2 using the ROC
  - Rate of Compression (ROC)=(Original Image size-Approx Image Size)/ Original Image size
- Getting the original Image back from step 1 and approximated image from step2

SVD compressed images for different values of k

Original Image

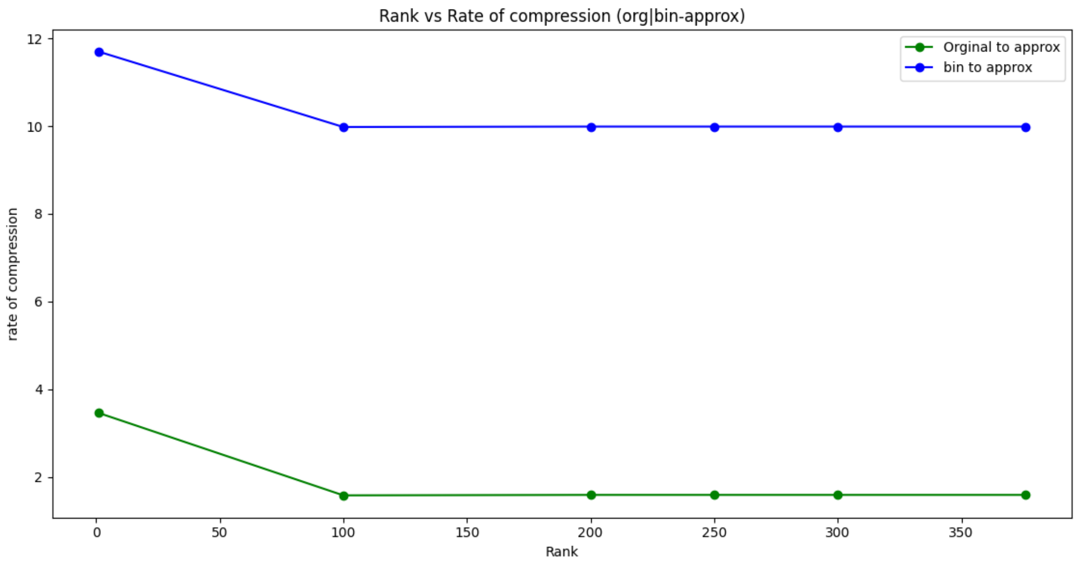
Rank 100

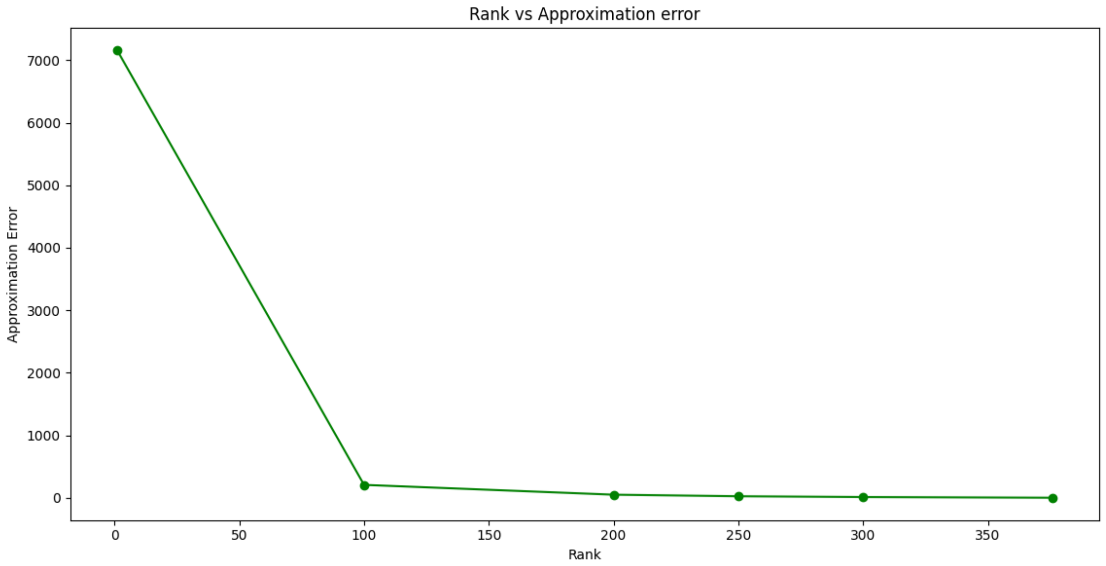
Rank 200

Rank 250

Rank 300

Observations





Observations

- Rates of Compression and Sizes for Different Test Cases with different ranks

Image	Dimensions	Ascii Size	Binary Size	Rank 5	ROC-5	Rank 10	ROC-10	Rank 20	ROC-20	Rank 40	ROC-40	Rank 80	ROC-80
Monalisa	250*360	360761	90000	6146	98.29%	12291	96.60%	24569	93.18%	49128	86.38%	98230	72.77%
Glass	320*480	470231	136960	7527	98.39%	15052	96.79%	30102	93.59%	60202	87.19%	120383	74.39%
Baseball	310*309	1245318	95790	6235	99.49%	12469	98.99%	24937	97.99%	49872	95.99%	99712	91.99%
hands	122*136	66433	16592	2625	96.04%	5246	92.10%	10483	84.22%	20953	68.45%	41872	36.97%

PCA

- Principal Component Analysis (PCA) is a dimension-reduction tool that can be used to reduce a large set of variables to a small set that still contains most of the information in the large set.
- Principal component analysis (PCA) is a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components.
- The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.
- PCA simply performs a coordinate rotation that aligns the transformed axes with the directions of maximum variance.
- PCA can help find underlying structure of the data. It selects a rotation such that most of the variability within the data s et is represented in the first few dimensions of the rotated data.
- PCA tries to find a lower dimensional surface so the sum of squares onto that surface is minimized.
- PCA tries to find the surface which has the minimum projection error.

Data Set- Iris Dataset

The considered Dataset has 4 attributes sepal and petal lengths and sepal and petal widths.
Three Classes Iris-setosa, Iris-virginica, Iris-versicolor

	sepal length	sepal width	petal length	petal width
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2

Data Processing

- The next step is to standardize the data

	sepal length	sepal width	petal length	petal width
0	-0.900681	1.032057	-1.341272	-1.312977
1	-1.143017	-0.124958	-1.341272	-1.312977
2	-1.385353	0.337848	-1.398138	-1.312977
3	-1.506521	0.106445	-1.284407	-1.312977
4	-1.021849	1.263460	-1.341272	-1.312977

Principal Components

	principal component 1	princial component 2
0	-2.264542	0.505704
1	-2.086426	-0.655405
2	-2.367950	-0.318477
3	-2.304197	-0.575368
4	-2.388777	0.674767

First principal component contains 72.77% of the variance and the second principal component contains 23.03% of the variance.
Together, the two components contain 95.80% of the information.

Results after adding targets