ADVANCED LEARNING FOR TEXT AND GRAPH DATA
[M2, MVA]

Maha ELBAYAD
`maha.elbayad@student.ecp.fr`

Lab 1 : Dimensionality Reduction

1 Unsupervised Dimensionality Reduction Techniques

1.1 SVD (Singular Value Decomposition)

We perform singular value decomposition on the image in Figure 1 (rank 480) and then reconstruct it using the k largest singular values ($k \in [10, 20, 50, 100, 200]$). The results are shown below with the reconstruction error values.

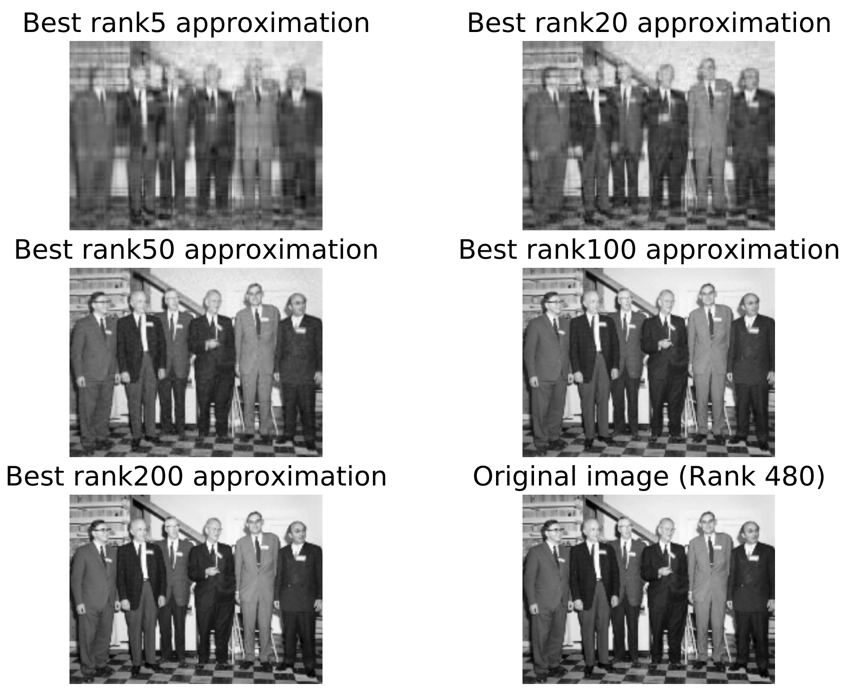
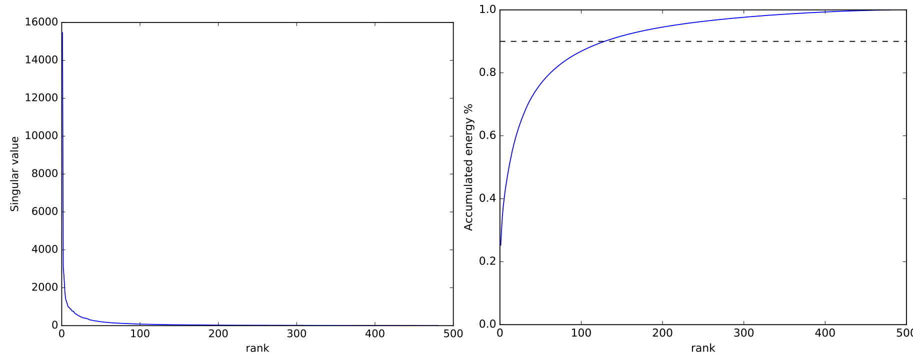


Figure 1: Original image and SVD reconstructions

k	10	20	50	100	200
Error	0.190	0.132	0.132	0.032	0.013

Table 1: Reconstruction errors

To choose the appropriate rank we plot the singular values $\sigma_1 \geq \sigma_2 ... \geq \sigma_n$ as well as the accumalted energy at each rank i.e $e_k = \frac{\sum_{j=1}^{k} \sigma_j}{\sum_{j=1}^{n} \sigma_j}$. We note that starting from $k = 100$ the restored image is less noisy as the lost information is minimal. To retain 90% of the singular values we may select $k^* = 130$ as the optimal rank.

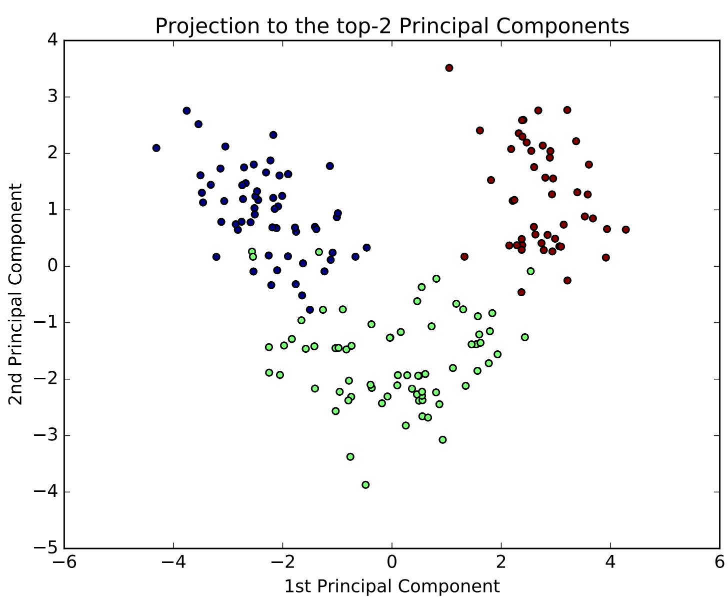


(a) Singular values (b) Accumulated energy

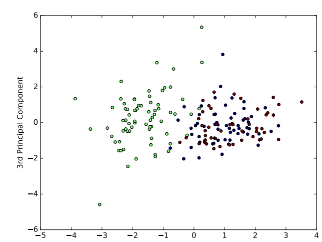
Figure 2

1.2 Principal Component Analysis (PCA)

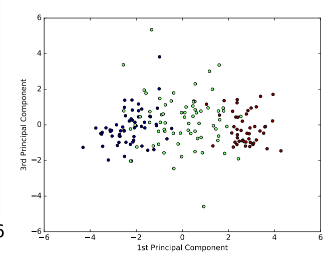
We perform PCA on the *Wine* dataset of size $(178, 13)$



(a) Data projection onto the plane (U_1, U_2)



(b) Plane (U_2, U_3)



(c) Plane (U_1, U_3)

We scatter the 3 classes in the first plane (U_1, U_2) and they are easily separable, thus we can reduce the data dimension from 13 to 2 efficiently.

2 Supervised Dimensionality Reduction and Classification with Linear Discriminant Analysis

We implement LDA as detailed in the assignement. When applied to the *Wine* 20 times we achieve an average accuracy of 98.4%

Average	Deviation	Min	Max
98.441%	1.204%	96.10%	100%