Unit 3

Dr Nakul Sharma
Assistant Professor (AI-DS)
VIIT Pune

Syllabus

Unit III:	Dimensionality Reduction Techniques

Overview of Dimensionality Reduction Techniques, Principal Component Analysis, Kernel Principal Component Analysis

Using Dimensionality Reduction Techniques for Finding an Eigen Portfolio, Curve Construction and Interest Rate Modelling, Bitcoin Trading: Enhancing Speed and Accuracy

Dimensionality Reduction

- When dealing with high dimensional data, it is often useful to reduce the dimensionality by projecting the data to a lower dimensional subspace which captures the "**essence**" of the data.
- This is called dimensionality reduction.
- Dimensionality reduction is a data preparation technique performed on data prior to modeling.
- It might be performed after data cleaning and data scaling and before training a predictive model.

- Dimensionality reduction: represent data with fewer dimensions.
- It has got following benefits
  - Easier learning
  - Visualization
  - discover “intrinsic dimensionality” of data

Principal Component Analysis (PCA)

- PCA is a useful way to summarize high-dimensional data (repeated observations of multiple variables).
- It also emphasizes consistent notation.
- The central ideas of PCA are **orthonormal coordinate systems**, the distinction between **variance** and **covariance**, and the possibility of choosing an **orthonormal basis** to **eliminate covariance**.

- PCA may be performed either by eigenvector analysis of the covariance matrix or by singular value decomposition of the original observation matrix.

Linear projection and reconstruction

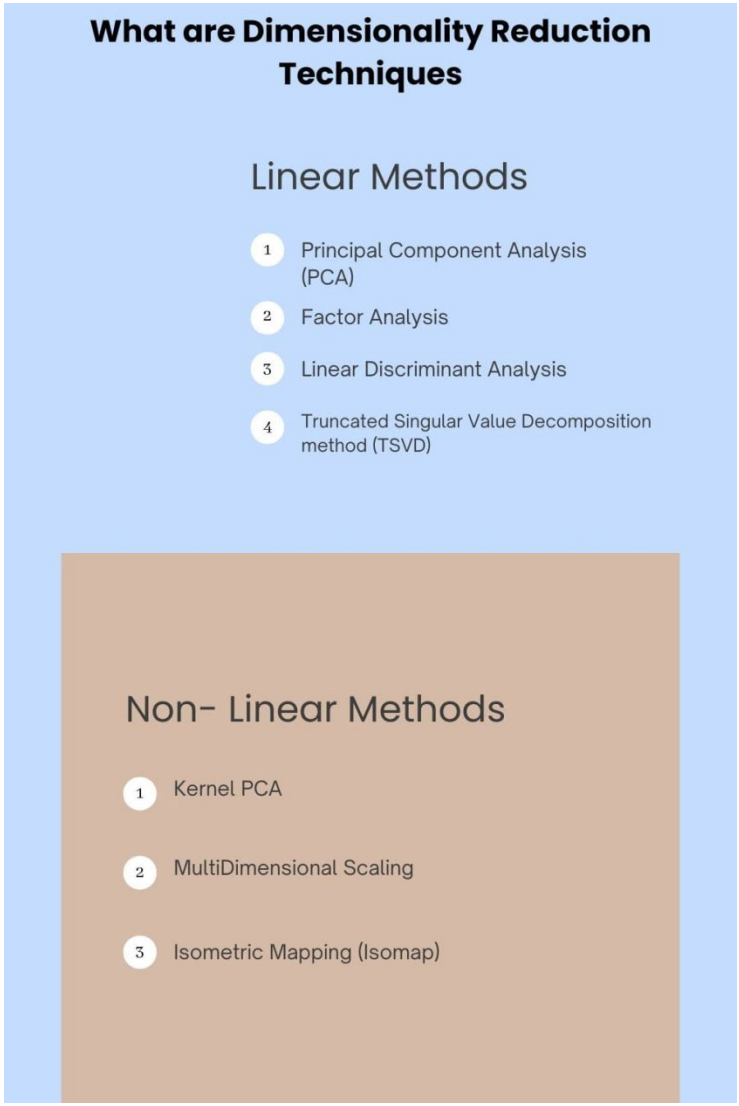
Project onto
1-dimension

Reconstruction:
Only knowing z, what
was (x_1,x_2)?

Principal component analysis (PCA) – Basic idea

- Project d-dimensional data into k-dimensional space while preserving as much information.
- Choose projection with minimum reconstruction error.

Dimentionality Reduction



References

- https://www.kdnuggets.com/2022/09/dimensionality-reduction-techniques-data-science.html