Machine Learning

Dimensionality Reduction using ICA

Satishkumar L. Varma

Department of Information Technology
SVKM's Dwarkadas J. Sanghvi College of Engineering, Vile Parle, Mumbai.
ORCID | Scopus | Google Scholar | Google Site | Website

Outline

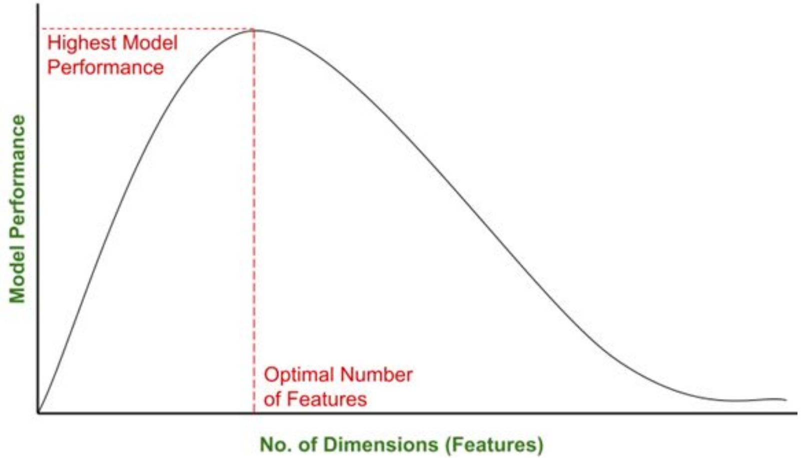
- Dimensionality Reduction
  - Dimensionality Reduction Techniques
    - Principal Components Analysis (Eigenvalues, Eigen vectors, Orthogonality)
    - Independent Component Analysis
    - Single Value Decomposition

Independent Component Analysis

- Introduction
- Data Science and ML helps to solve several complex regression and classification problems.
- However, the performance of all these models depends on the input dataset.
- So, It is important to provide optimal dataset to ML models.
- Large dataset leads to increased computational demands
- Large dataset leads to overfitting.
- If we provide large dataset (with a large no. of features/columns) to ML models
    - it gives rise to the problem of overfitting,
    - wherein the model starts getting influenced by outlier values and noise.
    - This is called the Curse of Dimensionality.

Dimensionality Reduction

- Dimensionality Reduction is a statistical/ML-based technique
- It helps to reduce the number of features in our dataset.
- It helps to obtain a dataset with an optimal number of dimensions.
- It is useful in Feature Extraction
  - By reduce the number of dimensions by mapping a higher dimensional feature space to a lower-dimensional feature space.
- Effect of the change in model performance with the increase in the number of dimensions of the dataset.
- The model performance is best only at an option dimension, beyond which it starts decreasing.
- Technique of Dimensionality Reduction (Feature Extraction)
  - Principal Component Analysis (PCA)
  - Independent Component Analysis (ICA)
  - Singular Value Decomposition (SVD)
  - Linear Discriminant Analysis (LDA)



Independent Component Analysis

- Independent Component Analysis (ICA) or Blind Signal Separation (BSS)
- ICA is used for the identification & separation of mixtures of sources with little prior information.
- It is a technique used to separate mixed signals into their independent, non-Gaussian components.
- It aims to find a linear transformation of data that maximizes statistical independence among the components.
- ICA is widely applied to isolate distinct sources from mixed signals:
  - Audio Processing
  - Image processing, and
  - Biomedical signal analysis
  - Medical data processing
  - Finance
  - Array processing (beamforming)
    - the process of linearly weighting and combining signals from array elements to form a desired spatial response pattern
  - Coding

Independent Component Analysis

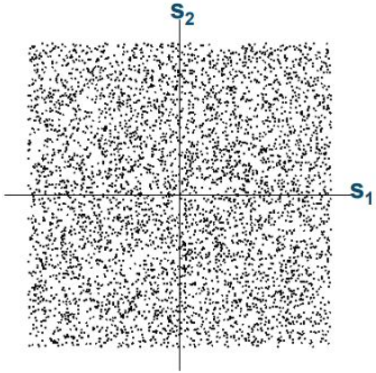
- Independent Component Analysis
  - PCA is based on the information given by the second
  - order statistics, whereas ICA goes up to high order statistics.
  - Therefore the result obtained by ICA is assumed to be more meaningful than the one gained by PCA.
  - However ICA better works on the data that have been already preprocessed by PCA.
- ICA operates under two key assumptions:
- 1. The source signals are statistically independent of each other.
- 2. The source signals have non-Gaussian distributions.
- These assumptions allow ICA to effectively separate mixed signals into independent components,
  - a task that traditional methods like PCA cannot achieve.

Independent Component Analysis

- Mathematical Representation of ICA
- The observed random vector is $X = (x_1, \ldots, x_\square)^T$
  - It represent the observed data with m components. random vector
- The hidden components are represented by $S = (s_1, \ldots, s_\square)^T$
  - where n is the number of hidden sources.
- Linear Static Transformation
  - The observed data X is transformed into hidden components S using a linear static transformation representation by the matrix W.
    - S = WX;
    - Where, W = transformation matrix.
  - Goal: Transform the observed data x such that the resulting hidden components are independent.
  - The independence is measured by some function $F(s_1, \ldots, s_\square)$
  - Task: Find the optimal W that maximizes the independence of the hidden components.

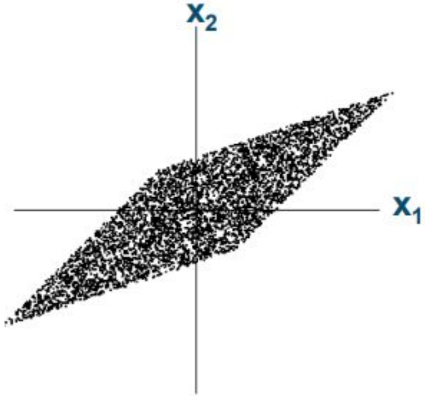
Independent Component Analysis

- Statistical Illustration of ICA
- Assume that the two individual components are mixed by the following mixing matrix,
- The mixed variables x can then be generated using the ICA model, x = As



$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad \text{i.e.} \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \Rightarrow \begin{cases} x_1 = 2s_1 + 3s_2 \\ x_2 = 2s_1 + 1s_2 \end{cases}$$

- The following shows joint distribution of the mixtures x1 and x2, and Notice anything interesting?

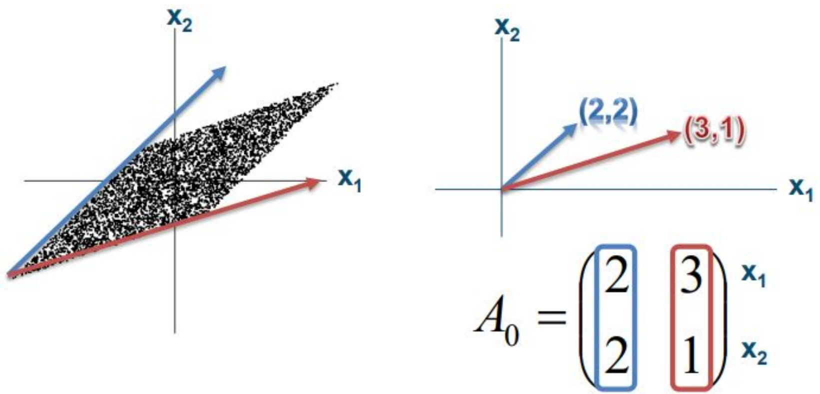


$$A_0 = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$$

Independent Component Analysis

- Statistical Illustration of ICA
- The edges of the parallelogram are the directions of the columns of A0.
- This means that we could, in principle, estimate the ICA model by first estimating the joint density x2, and then locating the edges.
- So, the problem seems to have a solution

$$A_0 = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$$



Independent Component Analysis

- This method works poorly in reality because it only works with variables that has uniform distributions.
- Moreover, it would be computationally quite complicated.
- What we need is a method that works for any distributions of the independent components, and works fast and reliably
- The goal of ICA is to find a linear mapping W such that the unmixed sequences u, are maximally statistically independent. $$\mathbf{u}(t) = \mathbf{W}\,\mathbf{x}(t) = \mathbf{W}\,\mathbf{A}\,\mathbf{s}(t)$$

Independent Component Analysis

- Advantages of ICA:
- ICA is a powerful tool for separating mixed signals into their independent components.
  - This is useful in a variety of applications, such as
    - signal processing,
    - image analysis, and
    - data compression.
- ICA is a non-parametric approach,
  - which means that it does not require assumptions about the underlying probability distribution of the data.
- ICA is an unsupervised learning technique,
  - which means that it can be applied to data without the need for labeled examples.
  - This makes it useful in situations where labeled data is not available.
- ICA can be used for feature extraction,
  - which means that it can identify important features in the data that can be used for other tasks,
    - such as classification.

Independent Component Analysis

- Disadvantages of Independent Component Analysis (ICA):
  - ICA assumes that the underlying sources are non-Gaussian, which may not always be true. If the underlying sources are Gaussian, ICA may not be effective.
  - ICA assumes that the sources are mixed linearly, which may not always be the case. If the sources are mixed nonlinearly, ICA may not be effective.
  - ICA can be computationally expensive, especially for large datasets. This can make it difficult to apply ICA to real-world problems.
  - ICA can suffer from convergence issues, which means that it may not always be able to find a solution. This can be a problem for complex datasets with many sources.

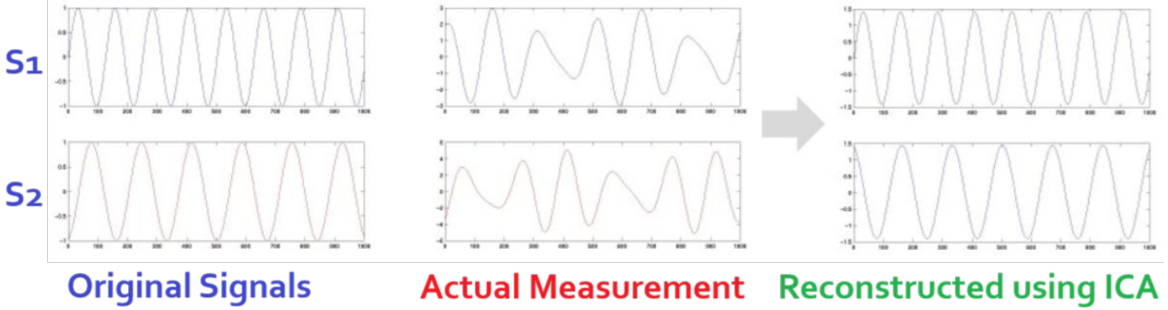
Independent Component Analysis

- Compare PCA and ICA

Principal Component Analysis	Independent Component Analysis
It reduces the dimensions to avoid the problem of overfitting.	It decomposes the mixed signal into its independent sources' signals.
It focuses on maximizing the variance.	It doesn't focus on the issue of variance among the data points.
It deals with the Principal Components.	It deals with the Independent Components.
It doesn't focus on the mutual independence of the components.	It focuses on the mutual independence of the components.
It focuses on the mutual orthogonality property of the principal components.	It doesn't focus on the mutual orthogonality of the components.

Independent Component Analysis

- ICA is unsupervised learning algorithms: We do not need to supervise the model before we can use it.
- The origin of ICA method comes from signal processing
  - Where we try to separate a multivariate signal into additive subcomponents.
  - Example



- Imagine two independent signals (S1 and S2) or variables represented as curves within the above image.
- The actual measurements, receive a data set that includes measurements of these two signals
  - which are unfortunately mixed into distinct linear combinations.
- ICA help to reconstruct/ recover the original signal from unknown signals by separating the mixed data.
- The ultimate aim is to reconstruct the data such that each dimension is mutually independent.

Independent Component Analysis

- ICA helps
  - To analyze a complex and highly correlated data set with large amount of information.
- To make this concept more tangible, let us understand ICA in three steps:
  - 1. Understand the fundamentals of ICA using cocktail parties problem (CPP)
  - 2. Perform the 3-step-ICA-algorithm, and
  - 3. Implement ICA for data analysis with R/Python using fastICA().

Independent Component Analysis

- 1. Understand the fundamentals of ICA using CPP
  - E.g. Cocktail party problem (CPP)

- ICA helps in the above such cases.
- ICA is a powerful technique in the field of data analysis.
- It allows to separate and identify the underlying independent sources in a multivariate data set.
- ICA is important because it provides a way to understand the hidden structure of a data set.
- ICA can help extract the most relevant (correlate) information from data, providing valuable insights.

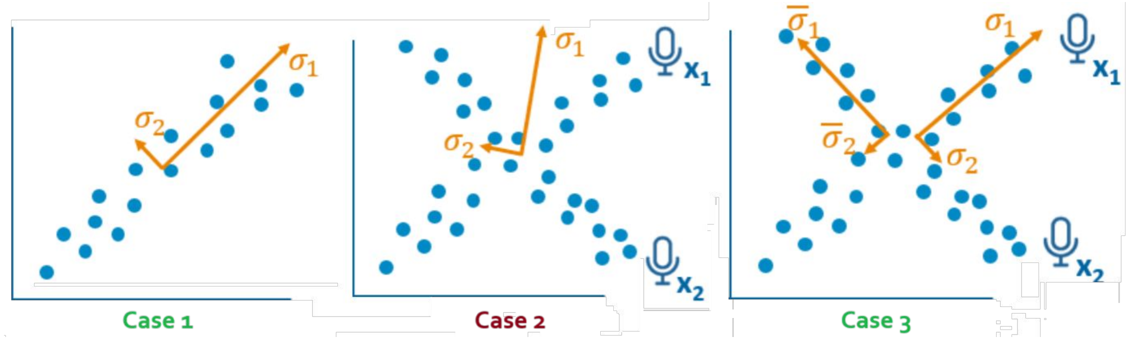
Independent Component Analysis

- 1. Understand the fundamentals of ICA using CPP
    - E.g. Cocktail party problem (CPP)

- Understand Mathematical Model (Framework)
    - 1st Measurement: X1 = $a_{11}*s_1 + a_{12}*s_2$
    - 2nd Measurement: X2 = $a_{21}*s_1 + a_{22}*s_2$
    - General Mathematical Model in vector notation: x = A * s
    - It can also be written as; x = Σa_i□s_i
    - Note: a_i□ are mixing coefficients or matrix which is assumed to be unknown.
- The starting point for ICA is the very simple assumption that the components s_i are statistically independent.
- The independent components are latent variables i.e. they cannot be directly observed.
- The measurements in vector X are actually the signals from vector s multiplied with
    - some mixing coefficients, represented in matrix A.
- All we observe is the random vector x_i and we must estimate A and s
- i.e To extract the full conversations (original signals), we need to solve this for given vector S.
- So, our goal is to estimate independent signals S such that S = A^{-1} * x

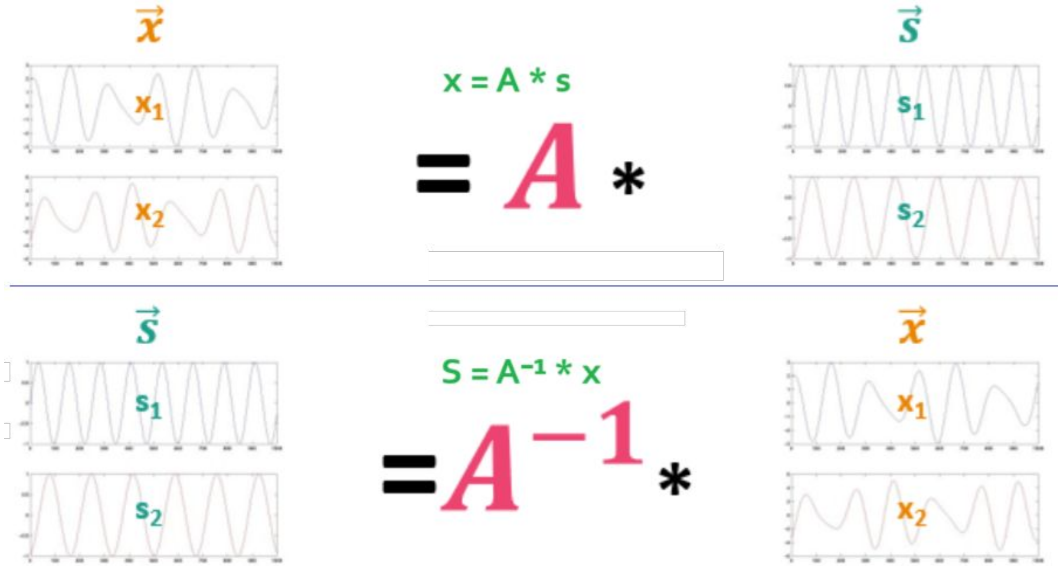
ICA versus PCA

- Recall the job of PCA
  - 1. Identify the principal direction of two related variables.
  - 2. By maximizing the variances using the eigenvector and eigenvalues of these variables,
    - PCA convert them into principal components.
- ICA is in some way related to PCA, and this assumption is not so wrong, But they differ in the last stage.
- Understand the different with an example using CPP.
  - The two measurements from two microphones have relationships that form something like a cross pattern.
  - If we were to apply PCA in CPP case, we would get the wrong results,
    - Case 1: PCA find main direction (principal components) in data,
    - Case 2: But PCA fails for data sets with more than one main (principal) direction.
    - Case 3: ICA solves this problem by focusing on ICs instead of PCs.



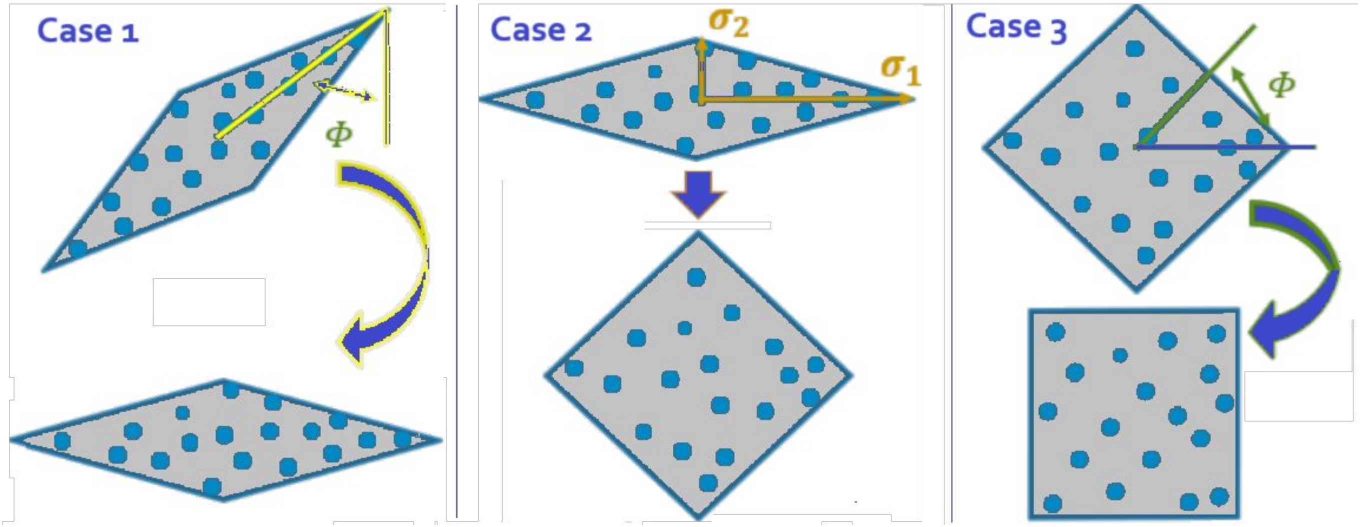
ICA versus PCA

- Recall the conceptual framework where matrix A is unknown.
  - All we observe is the random vector x_i and we must estimate A and s
    - i.e To extract the full conversations (original signals), we need to solve this for given vector S.
- So, our goal is to estimate independent signals S such that $S = A^{-1} * x$
  - To calculate the vector S, it is necessary to undertake inverse operations through a series of steps.
  - These sequential inverse operations comprise the three stages of the ICA algorithm.



Independent Component Analysis

- 2. Perform the 3-step-ICA-algorithm (Separation process -Splitting)
- Before lab demonstration in R or python, it is important to understand the three steps of the algorithm.
- The goal of the algorithm is to perform the multiplication of vector X with matrix A.
- Matrix A is comprised of three constituent parts,
  - which are the result of multiplicative interactions between the different factors:
    - $A = U^T \Sigma^{-1} V$
  - Case 1: U^T: Rotate
  - Case 2: Σ^{-1}: Stretch
  - Case 3: V: Rotate



Independent Component Analysis

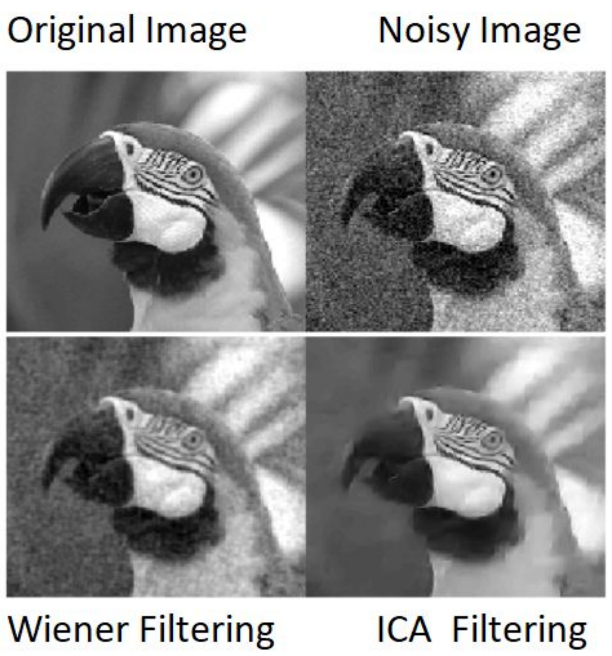
- 3. Implement ICA for data analysis with R/Python using fastICA()

Independent Component Analysis

- **Application domains of ICA**
  - Blind source separation (Bell & Sejnowski, Te won Lee, Girolami, Hyvarinen, etc.)
    - IEEE SIGNAL PROCESSING LETTERS, VOL. 6, NO. 4, APRIL 1999
  - Image denoising (Hyvarinen)
  - Medical signal processing – fMRI, ECG, EEG (Mackeig)
  - Modelling of the hippocampus and visual cortex (Lorincz, Hyvarinen)
  - Feature extraction, face recognition (Marni Bartlett)
  - Compression, redundancy reduction
  - Watermarking (D Lowe)
  - Clustering (Girolami, Kolenda)
  - Time series analysis (Back, Valpola)
  - Topic extraction (Kolenda, Bingham, Kaban)
  - Scientific Data Mining (Kaban, etc)

Independent Component Analysis

- **Application domains of ICA**
  - Image denoising



References

Text books:

1. Ethem Alpaydin, “Introduction to Machine Learning , 4th Edition, The MIT Press, 2020.
2. Peter Harrington, “Machine Learning in Action”, 1st Edition, Dreamtech Press, 2012."
3. Tom Mitchell, “Machine Learning”, 1st Edition, McGraw Hill, 2017.
4. Andreas C, Müller and Sarah Guido, “Introduction to Machine Learning with Python: A Guide for Data Scientists”, 1ed, O'reilly, 2016.
5. Kevin P. Murphy, “Machine Learning: A Probabilistic Perspective”, 1st Edition, MIT Press, 2012."

Reference Books:

6. Aurélien Géron, “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow”, 2nd Edition, Shroff/O'Reilly, 2019.
7. Witten Ian H., Eibe Frank, Mark A. Hall, and Christopher J. Pal., “Data Mining: Practical machine learning tools and techniques”, 1st Edition, Morgan Kaufmann, 2016.
8. Han, Kamber, “Data Mining Concepts and Techniques”, 3rd Edition, Morgan Kaufmann, 2012.
9. Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar, “Foundations of Machine Learning”, 1ed, MIT Press, 2012.
10. H. Dunham, “Data Mining: Introductory and Advanced Topics”, 1st Edition, Pearson Education, 2006.

Thank You.