In-Depth Explanation of Independent Component Analysis (FastICA)

Contents

1 Introduction

Independent Component Analysis (ICA) is a computational method used for separating a multivariate signal into additive, statistically independent components. This technique is widely used in various applications like audio signal processing, image processing, and more.

2 Definition of the Problem

The problem can be defined as follows: Given a set of observations X, we want to find an unmixing matrix W such that $S = W \times X$, where S are statistically independent signals. Mathematically, this can be represented as:

$$X = A \times S$$

Here, X is the observed data, A is the mixing matrix, and S are the original independent signals. We aim to find W, the unmixing matrix, such that:

$$S = W \times X$$

2.1 Derivation of Unmixing Matrix

The unmixing matrix W is essentially the inverse of the mixing matrix A, i.e., $W = A^{-1}$. This is derived from the equation $X = A \times S$ by multiplying both sides by A^{-1} to isolate S:

$$S = A^{-1} \times X$$

3 Data Preparation

3.1 Data Centering

Centering involves subtracting the mean from each variable (feature), making the mean of each variable zero. This is crucial because it simplifies the covariance matrix, which is essential for the whitening process.

$$X_{\text{centered}} = X - \mu$$

3.2 Data Whitening

Whitening transforms the data in such a way that its covariance matrix Σ becomes the identity matrix I. This is important because it decorrelates the features, which is a necessary condition for ICA.

$$Y = P \times X_{\text{centered}}$$

3.2.1 Mathematical Justification for Whitening

Starting from the definition of covariance, we have:

$$\mathrm{Cov}(Y) = \frac{1}{N} Y Y^T$$

Substituting $Y = P \times X_{\text{centered}}$ and simplifying, we get:

$$\begin{aligned}\mathrm{Cov}(Y) &= P \times \mathrm{Cov}(X) \times P^T \\ &= D^{-\frac{1}{2}} \times E^T \times \Sigma \times E \times D^{-\frac{1}{2}} \\ &= I\end{aligned}$$

4 Objective Function Optimization

The objective function in FastICA aims to maximize the non-Gaussianity of the output, which is measured using negentropy $J(y)$.

$$J(y) = H(y_{\text{gauss}}) - H(y)$$

4.1 Negentropy Approximation

Calculating $H(y)$ is computationally expensive. Therefore, we approximate negentropy $J(y)$ using the following equation:

$$J(y) = \left[\mathbb{E}\left\{\log\cosh(y)\right\} - \mathbb{E}\left\{\log\cosh(y_{\text{gauss}})\right\}\right]^2$$

4.2 Derivation of the Approximation

The approximation is derived from the Taylor series expansion of the cumulant generating function of y. The cumulant generating function is given by:

$$K(s) = \log \mathbb{E}\left\{e^{sy}\right\}$$

Expanding this in a Taylor series around $s = 0$ and comparing it with the Gaussian case, we arrive at the approximation for $J(y)$.

5 Iterative Optimization using Fixed Point Iteration

5.1 Fixed Point Iteration Scheme

The FastICA algorithm employs a fixed-point iteration scheme to find the unmixing matrix W.

$$w^+ = \mathbb{E}\{X \tanh(w^T X)\} - \mathbb{E}\{1 - \tanh^2(w^T X)\} w$$

5.2 Derivation of the Update Rule

The update rule is derived from the gradient ascent method applied to the objective function $J(w)$. Taking the gradient of $J(w)$ with respect to w, we get:

$$\frac{\partial J(w)}{\partial w} = \mathbb{E}\{X \tanh(w^T X)\} - \mathbb{E}\{1 - \tanh^2(w^T X)\}w$$

This gradient is then used to update w in each iteration, ensuring that the algorithm converges to a local maximum of $J(w)$.

6 Introduction

This document focuses on the objective function used in Independent Component Analysis (ICA). Specifically, we will discuss the concept of negentropy and why it is chosen as the objective function in ICA.

7 Objective Function in ICA

The objective function in ICA aims to maximize the non-Gaussianity of the output. One common measure of non-Gaussianity is negentropy $J(y)$.

$$J(y) = H(y_{\text{gauss}}) - H(y)$$

Here, $H(y)$ and $H(y_{\text{gauss}})$ represent the entropies of the random variables y and y_{gauss}, respectively.

7.1 Entropy and Negentropy

Entropy is a measure of the uncertainty or randomness associated with a random variable. For a continuous random variable y with probability density function $p(y)$, the differential entropy $H(y)$ is defined as:

$$H(y) = -\int p(y) \log p(y) \, dy$$

Negentropy $J(y)$ is used as a measure of non-Gaussianity and is defined as the difference between the entropies of a Gaussian random variable y_{gauss} and the random variable y:

$$J(y) = H(y_{\text{gauss}}) - H(y)$$

7.2 Why Negentropy?

The reason for using negentropy in ICA is rooted in the Central Limit Theorem, which states that the sum of independent random variables tends toward a Gaussian distribution. Since ICA aims to separate independent components, maximizing non-Gaussianity serves as a good objective function. This is because the independent components are the only non-Gaussian signals that contribute to the observed data.

8 Approximating Negentropy

Calculating $H(y)$ is computationally expensive. Therefore, we approximate negentropy $J(y)$ using the following equation:

$$J(y) \approx \mathbb{E}\{\log\cosh(y)\} - \mathbb{E}\{\log\cosh(y_{\text{gauss}})\}$$

8.1 Why Log Cosh?

The function $\log\cosh(y)$ is used because it has properties similar to entropy and is computationally easier to handle. This approximation is particularly useful because the exact computation of $H(y)$ involves calculating an integral over the probability density function, which can be computationally expensive.

9 Conclusion

The objective function in ICA aims to maximize the non-Gaussianity of the separated components, and negentropy serves as an effective measure for this purpose. Due to computational considerations, negentropy is often approximated using functions like $\log\cosh(y)$.