

A Tutorial on Independent Component Analysis

矩阵分析与应用 2022

田松松 彭方捷 黄河 杨侯哲

2022 年 5 月 22 日

中国科学院大学

Table of contents

1. 问题来源
2. 模型建立
3. ICA 求解策略
4. 观测数据的协方差
5. 白化处理
6. 条件独立性
7. 一种信息论的解法
8. 例子
9. 总结

问题来源

在对现实数据进行测量的时候，有时数据是多个独立分量的组合，例如当研究者录下声音进行统计分析的时候，录下的声音可能来自多个音源，比如街道上录下的声音就含有谈话声、脚步声、汽车声等成分。研究者希望从原始数据中将这些独立成分提取出来。

这个任务被称为盲源分离 (BSS, Blind Source Separation)。这个问题一般而言是难解的，然而在部分特殊情形下，这个问题近期得到了解决。这就是独立成分分析 (ICA, Independent Component Analysis)。

示例

一个常见的例子是声音数据，因为根据物理学，声音是线性叠加的。让我们考虑一个简单的模型：有一场派对，派对中有两个音源：音乐和人声，场上有两个录音机，录音机记录的声音是来自两个音源的声音的线性叠加，其权重根据录音机和音源距离而变化。

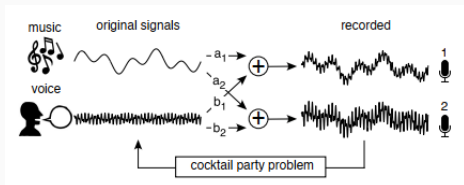


图 1: 模型示意

示例

设音乐和人声的响度分别为 s_1, s_2 ，两个录音机接收的声音为 x_1, x_2 。由声音的线性叠加，我们有

$$x_1 = a_1 s_1 + b_1 s_2$$

$$x_2 = a_2 s_1 + b_2 s_2$$

即

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

示例

下图是数据点的示例，其中横纵坐标分别表示声音记录在两个录音机上的振幅。单个音源的数据记录分布在一条直线上，因为音源到两个录音机的距离是固定的。右图为两个音源的叠加，颜色表示在这个点来自哪个音源的声音占据主导（颜色只为示意，不在原始数据中）。

在这个示例中，由于两个音源的声音分布均集中在中心，叠加数据也集中在中心，远离中心的数据则主要分布在两条带上。

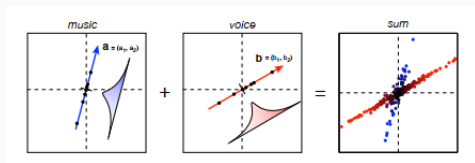


图 2: 数据示例

左图为 $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} s_1$ ，中图为 $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} s_2$

示例

下图中的第二行是另一个示例。此时音源 2 的数据分布呈现双峰，使得观测数据聚集在两个区域。

下图中的第三行则是一个垂直的正态分布的示例。

右数第二幅图中的红线和蓝线表示 ICA 得到的两个独立成分，最右图中的绿线表示最大方差的方向。仅在三幅图中，独立成分分析得到和主成分分析一样的结果。

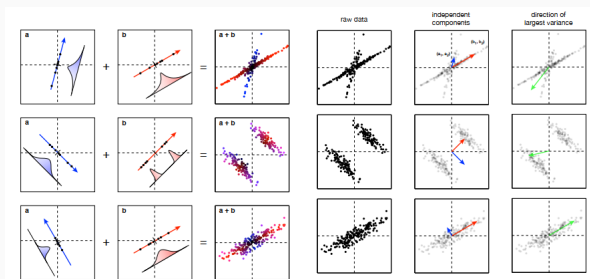


图 3: 数据示例

模型建立

模型建立

- 对于多元数据 x , 我们假设每个样本都是从未知分布 $P(x)$ 中随机抽取的。不失一般性的, 我们认为 x 不止一个维度。
- 对于盲源 s , 我们假设其每个源 s_i 是统计独立的, 即

$$\forall i \neq j, p(s_i, s_j) = p_i(s_i) p_j(s_j)$$

- ICA 思想的关键是假设观测数据 x 是盲源 s 的线性混合

$$x = As \tag{1}$$

其中 A 是混合源分量的某个未知可逆方阵。在图1的例子中

$$A = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$$

- ICA 的目标是找到混合矩阵 A (更具体地说, A 的逆矩阵), 以便从观测数据 x 中恢复原始信号 A 。

从以上分析中知道，只需要构建新的矩阵 W ，使得观测数据 x 的线性变换是盲源 s 的估计

$$\hat{s} = Wx \quad (2)$$

这似乎是一个很难解决的问题，仅仅通过观察矩阵 A 和 s 的乘积 x 来求解它们。这类似于仅通过观察 $c = a \times b$ 来求解 a 和 b 。在数学上这被称之为欠约束问题，因为未知数的个数超过了已知量的个数。

ICA 求解策略

分而治之

分而治之提供了解决 ICA 问题的一个思路。不试图同时求解 s 和 A ，而是先求解矩阵 A 。求解矩阵 A 时也不是一次性求出，而是通过将 A 分解成更简单的部分，每次只求解 A 的一小块。

奇异值分解 (SVD) 在线性代数与自动控制领域应用很广泛，它提供了一种简单有效的矩阵分解方法。对于任何矩阵，它的奇异值分解表示

$$A = U\Sigma V^T$$

其中 U 是一个 $m \times m$ 的矩阵， Σ 是一个 $m \times n$ 的矩阵，除了主对角线上的元素以外全为 0，主对角线上的每个元素都称为奇异值， V 是一个 $n \times n$ 的矩阵。 U 和 V 都是酉矩阵，即满足 $U^T U = I$ ， $V^T V = I$ 。

SVD 的几何意义

任何矩阵都被分解为三个更简单的线性变换：旋转 V 、沿轴拉伸 Σ 和第二次旋转 U ， SVD 分解得到的每个矩阵都比原矩阵更简单，每个矩阵包含的参数更少，并且每个矩阵的求逆也很容易。 U 和 V 是旋转矩阵（或正交矩阵）， Σ 是具有实非负值的对角矩阵。

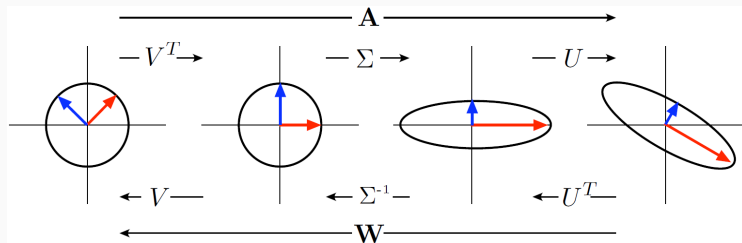


图 4: SVD 几何意义

图中红色和蓝色箭头是对应于矩阵 V 的列向量（即 A 的行空间的基）。注意基底在每次连续操作中是如何旋转、拉伸和转动的。所有三个线性变换的组合等价于矩阵 A 的线性变换，矩阵 A 的逆矩阵定义为 $W = V\Sigma^{-1}U^T$ ，即以相反的顺序执行每个线性变换的逆变换。

我们通过单独求解 SVD 分解的每一部分来估计 A，即其逆 W

$$W = A^{-1} = V\Sigma^{-1}U^T \quad (3)$$

该公式的主要依据是旋转矩阵的逆矩阵是其转置矩阵。同时，因为假设 A 是可逆的，因此 Σ^{-1} 存在且有意义。

下面主要通过两个步骤来求解混矩阵 W

- 1 计算观测数据 x 的协方差，从而来计算 U 和 Σ 。
- 2 从盲源 s 的独立性假设出发求解 V。

观测数据的协方差

协方差矩阵

本节将探索给定假设下数据的协方差，并在此过程中推导出构成矩阵 W ($W = V\Sigma^{-1}U^T$) 的三个矩阵中的两个。数据的协方差提供了一个很好的思路，因为协方差矩阵衡量了线性模型不同分量之间的所有相关性。

- 协方差矩阵是方阵且是对称矩阵。
- 其中第 i 行第 j 列的值是随机变量 x_i 和 x_j 之间的协方差，该值反映了随机变量 x_i 和 x_j 之间的二阶统计特性。
- 为简单起见，假设数据的均值为零，或每个维度都已经中心化处理。

协方差矩阵都可以通过随机向量之间的外积的数学期望来定义，即 $\text{cov} = E\langle \mathbf{x}\mathbf{x}^T \rangle$ ，为了恢复矩阵 Σ 和 U ，增添一个额外假设：

假设盲源 s 的协方差是白化的 $\iff \langle \mathbf{s}\mathbf{s}^T \rangle = \mathbf{I}$

其中 \mathbf{I} 是单位矩阵。

协方差矩阵

将式 (1) 代入随机变量的外积 $\langle \mathbf{x}\mathbf{x}^T \rangle$ 可得

$$\begin{aligned}\langle \mathbf{x}\mathbf{x}^T \rangle &= \langle (\mathbf{A}\mathbf{s})(\mathbf{A}\mathbf{s})^T \rangle \\ &= \langle (\mathbf{U}\Sigma\mathbf{V}^T\mathbf{s})(\mathbf{U}\Sigma\mathbf{V}^T\mathbf{s})^T \rangle \\ &= \mathbf{U}\Sigma\mathbf{V}^T \langle \mathbf{s}\mathbf{s}^T \rangle \mathbf{V}\Sigma\mathbf{U}^T\end{aligned}$$

根据关于盲源 \mathbf{s} 的白化假设 ($\langle \mathbf{s}\mathbf{s}^T \rangle = \mathbf{I}$) 和正交矩阵的性质 ($\mathbf{V}^T = \mathbf{V}^{-1}$), 化简得到最终表达式

$$\langle \mathbf{x}\mathbf{x}^T \rangle = \mathbf{U}\Sigma^2\mathbf{U}^T \quad (4)$$

观察上式发现, 通过合理的假设, 数据的协方差与盲源 \mathbf{s} 和 \mathbf{V} 均无关!

等式 (4) 的特别之处在于它用两个正交矩阵 \mathbf{U} 及其之间的对角矩阵 Σ^2 表示数据的协方差。

显然， $\langle \mathbf{x}\mathbf{x}^T \rangle$ 为对称矩阵，而由矩阵论相关知识我们可知，任何实对称矩阵（包括协方差矩阵）都通过其特征向量正交对角化。构造矩阵 E ，其列是 $\langle \mathbf{x}\mathbf{x}^T \rangle$ 的特征向量，可证明

$$\langle \mathbf{x}\mathbf{x}^T \rangle = EDE^T \quad (5)$$

其中 D 是对角矩阵，其对角线上的元素为对应的特征值， E 是对应的特征向量构成的正交矩阵（数据协方差的特征向量形成一个正交基）。

比较式（4）和式（5），两个方程都表明数据的协方差可以通过正交矩阵对角化。式（4）提供了基于 ICA 基本假设的分解。式（5）提供了完全基于对称矩阵性质的特征分解。

特征分解

使用其特征向量对是角化对称矩阵的唯一方式，即没有其他基可以对角化对称矩阵。

因此，如果关于 ICA 的假设是正确的，那么目前已经确定了矩阵 A 的部分解：U 是数据协方差的特征向量构成的矩阵， Σ 是对角矩阵，对角线上是特征值的平方根。

通过合理的构造，目前我们已经确定了式 (3) 中新矩阵 W 中的两个矩阵，即

$$W = V\Sigma^{-1}U^T = VD^{-\frac{1}{2}}E^T$$

其中，D 和 E 是由观测数据 x 的协方差矩阵的特征值和特征向量构成的，V 是唯一未知的旋转矩阵。

白化处理

白化的几何解释

上面对 *ICA* 问题的求解过程与信号处理中常见的操作类似，称为白化。白化是一种去除数据集中所有线性相关性（即二阶相关性）并沿所有维度归一化方差的操作。通俗地说，这种操作被称为球化数据，因为直观上，白化将数据映射成球对称分布。图（5）给出了两个变换 $D^{-\frac{1}{2}}E^T$ 的几何解释：

- 1 首先旋转数据以沿笛卡尔基对齐协方差的特征向量。通过乘以 E^T 执行这种旋转，并使数据去相关，即去除线性相关性。数学上，去相关意味着变换后数据的协方差被对角化。

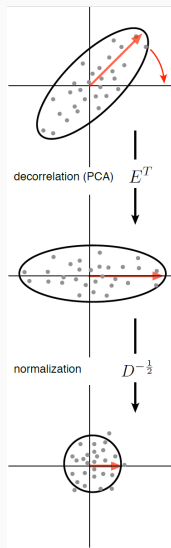


图 5: 白化的几何意义

针对 ICA 模型有

$$\left\langle (E^T x) (E^T x)^T \right\rangle = D$$

其中 D 是由特征值构成的对角矩阵，对角线上的每个元素都是数据协方差的特征值，衡量了每个维度方差的大小。

这个操作有一个被广泛熟知的名字，叫做主成分分析 (PCA)。数据协方差的特征向量是对应的主成分。将数据集投影到主成分上消除了线性相关性，并提供了降维策略（通过选择性地去除低方差的维度）。

- 2 第二步操作通过乘以 $D^{-\frac{1}{2}}$ 来归一化各维的方差，从而确保了所有维度都以标准单位向量表示，不存在主方向，数据均是旋转对称的，类似于球体。

可以看出，通过白化处理后，ICA 要做的就是寻找一组正交旋转矩阵 V 。可做如下定义

$$x_w = \left(D^{-\frac{1}{2}} E^T\right) x$$

其中 x_w 是经观测数据白化处理得到的，使得 $\langle x_w x_w^T \rangle = I$ 。将上式代入式 (2) 和式 (3)，可将 ICA 问题简化为求解 $\hat{s} = Vx_w$ 。而最终的问题为寻找旋转矩阵 V 。

ICA 的简化模型为恢复盲源 \hat{s} 提供了额外思路。图5强调了白化数据 x_w 是旋转对称的，因此，白化数据旋转后得到的 \hat{s} 也肯定是白化的（解 $\hat{s} = Vx_w$ ）。这与在第六节中对 s 的假设是一致的，同时这也说明存在着多个白化滤波器，包括 $D^{-\frac{1}{2}} E^T$ 和 $V \left(D^{-\frac{1}{2}} E^T\right)$ 。

条件独立性

根据前文，ICA 方法的目的是找到线性变换的矩阵 W ，用它来将已知数据恢复为线性混合的数据源 s ，即 $s = Wx$ 。

ICA 方法假设已知数据 x 是数据源 s 的线性混合，即 $x = As$ ，其中 A, s 为未知量，通过对矩阵 A 的分解和对数据的白化，上述问题已经转化为找到一个旋转矩阵 V 得到近似的数据源 $\hat{s} = Vx_w$ 。

之前提到过，协方差矩阵测量了所有变量对之间基于二阶相关性的线性相关程度，而白化数据移除了变量间的二阶相关性，因此想要找到旋转矩阵 V 需要通过其他的手段来测量相关性。

因此我们将要利用条件独立性来确定 V 。

条件独立的特性

条件独立是随机变量间最严格的测量独立性的手段，它不仅要求变量间不具有二阶相关性，更进一步的，它要求变量间不具有任意更高阶的相关性。

举例来说，如果随机变量 a, b 之间条件独立，那么可知 $P(a, b) = P(a)P(b)$ ，联合概率也就被分解开来。

用在 ICA 中，我们已经假设所有的数据源相互条件独立，那么也就可知

$$P(s) = \prod_i P(s_i)$$

这也表明了数据源的联合分布 $P(s)$ 是一个特殊的分布族**阶乘分布**，因为联合分布的概率是每个数据源分布 $P(s_i)$ 的乘积。

寻找旋转矩阵

已知 ICA 要寻找的旋转矩阵 V 求得的近似数据源应该是满足条件独立性的，即 $P(\hat{s}) = \prod_i P(\hat{s}_i)$ 。

之前已经说过，白化过程移除了数据间的二阶相似性，那么旋转矩阵就必须移除掉数据间剩余的所有高阶相似性，这看起来是个不可能完成的任务，但是如果我们的假设是正确的，即 $x = As$ ，那么这个过程是可行的并且 \hat{s} 将是条件独立的。

因此可以定义一个**对比函数**，用来衡量数据间的高阶相似性，或者等价地说，衡量得到的近似数据源 \hat{s} 距离条件独立有多近，通过找到函数的最值，就可以找到最理想的近似数据源。

一种信息论的解法

为了找到旋转矩阵 V 使近似数据源 \hat{s} 条件独立，我们需要用到信息论中的知识。

在信息论中，为了衡量近似数据源 \hat{s} 的条件独立性，存在许多可能的对比函数，例如对相关性的严格统计度量，对这种度量的近似，以及一些优雅独特的猜测方法。

作为基础，这里主要使用信息论中一种自然产生的测量方法来测量一个分布距离条件独立的距离，也就是**互信息**。互信息测量了两个变量与条件独立的背离程度，而多重信息，也就是互信息的推广，测量了多个变量间与条件独立背离的程度：

$$I(y) = \int P(y) \log_2 \frac{P(y)}{\prod_i P(y_i)} dy \quad (6)$$

在信息论中，任何一个事件都有对应的信息量，通俗地说，事件发生的概率越低，那么这个事件包含的信息量越大，这和我们直观上的认知也是一致的。

而为了具体的衡量信息量，信息论中提出了熵这一概念，它量化的表示了随机变量中包含的“信息量”。它被定义为

$H(X) = -\sum_{x \in X} p(x) \log p(x)$ 。而互信息，就表示对于两个随机变量，如果我们知道了事实 X 后，对于 Y 的熵减少了多少，或者说，我们“知道”了多少关于 Y 的信息。

互信息的形式定义为

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (7)$$

那么当两变量条件独立时，也就意味着它们不存在相关性，知道任何一个变量并不会减少另一变量的熵，也就表示此时的互信息为 0。

当然，互信息的形式可以推导为和熵直接相关的形式

$$\begin{aligned} I(X, Y) &= \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x)p(y|x)}{p(x)p(y)} \\ &= \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(y|x)}{p(y)} \\ &= \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(y|x) - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(y) \\ &= - \sum_{y \in Y} p(y) \log p(y) - [- \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(y|x)] \\ &= H(X) - H(X|Y) \end{aligned}$$

同样的，互信息的扩展多重信息也可以推导为类似形式。

互信息的直观表示

上面提到的互信息与熵的关系可以通过下图直观的了解

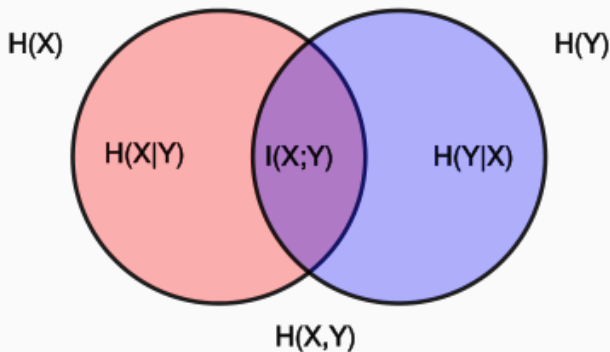


图 6: 互信息与熵的关系

多重信息是非负的，而当它等于 0 时只可能是所有的变量都条件独立，也就是当 $P(y) = \prod_i P(y_i)$ ，那么上述公式可知 $\log(1) = 0, I(y) = 0$ 。

因此我们当前的目标非常简单，找到一个旋转矩阵 V 使得多重信息 $I(\hat{s}) = 0$ ，其中 $\hat{s} = Vx_w$ ，如果能找到这样的一个旋转矩阵，那么 \hat{s} 是条件独立的，更进一步的， $W = VD^{-\frac{1}{2}}E^T$ 是 ICA 的解，可以用来估计数据潜在的来源。

两个随机变量的情况

多重信息当近似数据源 \hat{s} 条件独立时最小，考虑如下图的情况

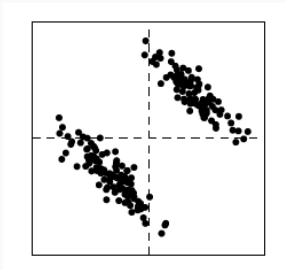


图 7: 二维数据样例

当数据为二维时，旋转矩阵 V 有如下形式

$$V = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

旋转角度

上述旋转矩阵实际只有一个自由变量，也就是旋转角度 θ 。对于这种情况我们可以计算出对所有的 θ 对应的多重信息 \hat{s} ，实际结果如下图

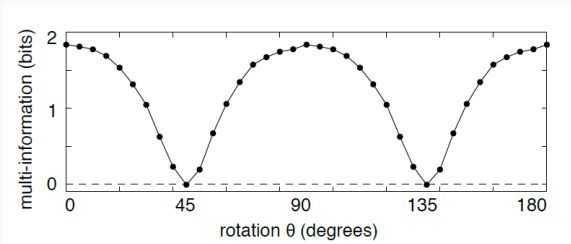


图 8: 多重信息随旋转角度的变化

从上图可以看到，当旋转的角度为 45° 时，多重信息为 0，这表示旋转到这个角度时恢复的分布间是条件独立的，也就代表对应的旋转矩阵 V 是一个二维的 45° 旋转矩阵。

旋转数据的效果

表面看起来这样的过程很抽象，我们可以把这个过程可视化出来，对于旋转了不同角度的近似数据源 \hat{s} 如下图

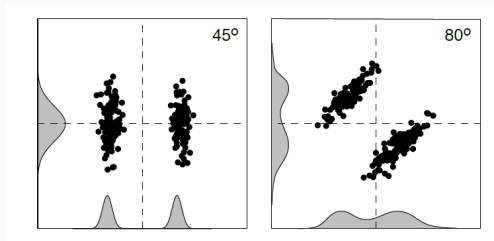


图 9: 不同旋转角度下的近似数据源 \hat{s}

可以看出，当旋转的角度为 45° 时，近似数据源 \hat{s} 在 x 轴与 y 轴上分别为双峰和单峰的高斯分布，两分布间条件独立。另外需要说明的时，旋转角度不是唯一的，多旋转 90° 后仍然可以使多信息达到最小。

简化形式

最小化多重信息在现实情况下是比较困难的，但这个过程可以简化，优化的简化过程揭示了与 ICA 其他形式之间的重要关系。

多重信息是一个分布关于熵 $H[\cdot]$ 的函数。熵的定义为 $H[\cdot] = -\int P(y)\log_2 P(y)dy$ ，它测量的是一个分布 $P(y)$ 中的不确定性。而多重信息则为边缘分布熵的和与联合分布的熵的差值，即 $I(y) = \sum_i H[y_i] - H[y]$ 。因此近似数据源 \hat{s} 的多重信息为

$$\begin{aligned} I(\hat{s}) &= \sum_i H[(Vx_w)_i] - H[Vx_w] \\ &= \sum_i H[(Vx_w)_i] - (H[x_w] + \log_2 |V|) \end{aligned}$$

其中 (Vx_w) 表示近似数据源 \hat{s} 的第 i 个元素，上述公式还应用了线性变换下概率密度的熵的变换形式。

线性变换下概率密度熵的变换

假定有随机变量 $y = Ax$ ，那么根据概率论中随机变量的变量变换可知 $p_y(y) = \frac{1}{|A|} p_x(x)$ 。而在随机变量的熵可以理解为

$$\begin{aligned} H[y] &= -E[\log p_y(y)] \\ &= -E[\log \frac{1}{|A|} p_x(x)] \\ &= -E[\log p_x(x)] + E[\log |A|] \\ &= H[x] + \log |A| \end{aligned}$$

进一步化简

由于一个旋转矩阵已知它的行列式为 1，因此可知 $\log_2|V| = 0$ 。更进一步的，由于我们实际的目标是找到旋转矩阵 V 而非找到上述多重信息的最小值，因此我们可以发现 $H[x_w]$ 是与 V 无关的常量，去掉无关量后我们可以得到

$$V = \arg \min_V \sum_i H[(Vx_w)_i] \quad (8)$$

优化过程简化为找到使近似数据源 \hat{s} 的边缘分布熵之和最小的旋转矩阵，该旋转矩阵 V 最大化了近似数据源 \hat{s} 之间的条件独立性。

对优化过程的不同理解

首先需要说明的是，根据一个有限数据集计算熵是一件很困难的事，需要小心，因此有很多的 ICA 优化策略是关注于如何求得公式8的近似解。

公式8的形式可以有多种理解，这反映了关于 ICA 方法不同但等价的解释。一种理解为，公式8找到的旋转最大化了经过变换的数据的“非高斯性”。另一种解释为，公式8等价于在假设观测数据来自于条件独立分布的情况下，找出使观测数据的对数似然最大化的旋转矩阵。这两种解释都为其他作者推导公式8提供了起点。

非高斯性的解释

一个分布与高斯分布的距离通常用负熵来衡量，一个分布的负熵与一个分布与相同均值和方差的高斯分布的 KL 散度相等，因此它被称为非高斯性。

KL 散度是衡量两个分布差别的非对称性的度量，定义如

$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} dx$ 。另一方面，KL 散度和互信息也同样有着关联 $I(X, Y) = D_{KL}(P(X, Y)||P(X)P(Y))$ 。

因此最大化非高斯性也就是最大化对应的 KL 散度，等同于最小化边缘分布的熵。

经过上述过程，我们得到了一个优化方法允许我们估计矩阵 V 从而近似的重建条件独立的原始数据源 $\hat{s} = Wx$, 而 W^{-1} 的各列分别代表了数据中的独立分量。

例子

Fourth-Order Blind Identification(FOBI)

FOBI 法可能是独立成分分析的第一种解法，在 1989 年被 Cardoso 提出，使用四阶距进行盲分离。FOBI 分为两步：白化和分解二阶加权协方差阵 $\Omega = E[|X_w|^2 X_w X_w^T]$ 。进行白化排除相关性后，独立信号源可以通过加权协方差矩阵的特征向量进行辨识。

Ω 是实对称矩阵，可被么正对角化。下面求 Ω 的谱分解。由 $\hat{S} = VX_w$ 有，

$$\Omega = E[|s|^2 V s s^T V^T] = V E[|s|^2 s s^T] V^T, |s|^2 = \sum_{k=1}^N s_k^2$$

对于 $E[|s|^2 s s^T]$ ，在对角线上时，有 $i = j$ ，当 $k = i$ 时，有 $E(s_i^4)$ ，当 $k \neq i$ ，有 $E(s_k^2)E(s_i^2) = 1$ ，有 $n - 1$ 项。在非对角线上有 $E[|s|^2 s_i s_j^T] = 0$ 。所以 $E[|s|^2 s s^T]$ 可化为 $\text{Diag}[E(s_i^4) + N - 1]$ 。

随机生成数据的 ICA

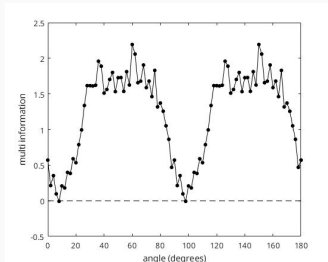


图 10: 多重信息随旋转的变化

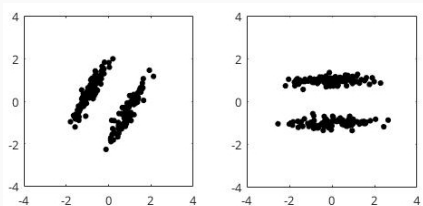


图 11: 原数据和旋转后数据

论文的结尾提供了一个例子，进行以下步骤：

(1) 随机生成数据; (2) 白化，计算不同角度下的多重信息; (3) 基于 FOBI 解 ICA。

运行操作系统：Linux Deepin 20

开发语言: matlab

开发工具：matlab2016

ICA 用于音频和图片处理



图 12: 原图



图 13: 源 1



图 14: 源 2

音频处理: 混合音频 源 1 源 2

运行操作系统: Linux Deepin 20

开发语言: python2

开发工具: pycharm

总结

ICA 的目标是从若干独立成分的线性叠加中将每个成分分离出来。其步骤可以归结为：

- 减去数据的均值，将数据平移到以原点为中心
- 通过旋转和伸缩将数据的协方差阵变为单位阵（此过程称为白化）
- 通过旋转使数据的各个分量之间互信息最小（从而接近独立）

在线性叠加的假设下，ICA 作为一种分离独立成分的方法，不需要假定正态分布，也不需要假定各个成分互相垂直，因而应用比 PCA 更加广泛。

在分解 $W = V\Sigma^{-1}U^T$ 中， U 和 Σ 消除了变量间的二阶相关性，而 V 则尝试借助信息论中的方法消除变量间的高阶相关性。由于多元正态分布中变量之间只有二阶相关，PCA 足以提取所有信息，此时 PCA 和 ICA 等效。注意到 U 和 Σ 可以解析求解， V 则没有解析解，只能通过一些数值分析的方式求解，这是进行 ICA 的一个难点。

本文的 ICA 尚有可以继续改进的地方。下图是一些 ICA 无法处理的例子。第一幅图对应于独立分量数高于观测值维数的情形，此时 A 不是方阵，因而不可逆。第二幅图对应于数据分布于非线性流形的情形，第三幅图对应于数据间的叠加非线性情形，因而 ICA 无法处理。这些都是关于 ICA 新的研究方向。

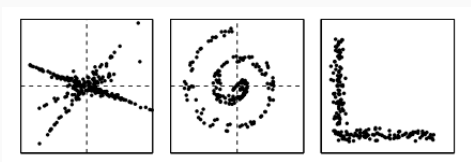


图 15: 一些 ICA 无法处理的例子