Review of Independent Component Analysis

Element of Statistical Learning Book, Chapter 14

Hanchao Zhang

November 9, 2021

1 Latent Variables an Factor Analysis

The singular decomposition

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T \tag{1}$$

We can write $\boldsymbol{S} = \sqrt{N}\boldsymbol{U}$ and $\boldsymbol{A}^T = \frac{\boldsymbol{D}\boldsymbol{V}^T}{\sqrt{N}}$
and we have

$$\boldsymbol{X} = \boldsymbol{S}\boldsymbol{A}^T = \sqrt{N}\boldsymbol{U}\frac{\boldsymbol{D}\boldsymbol{V}^T}{\sqrt{N}} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T \tag{2}$$

Where $\boldsymbol{S}$ and $\boldsymbol{X}$ have mean 0, and $\boldsymbol{U}$ is an orthogonal matrix. We can interpret the SVD or the corresponding principal component analysis as an estimate of a latent variable model.

$$\begin{aligned} X_1 &= a_{11}S_1 + a_{12}S_2 + \cdots + a_{1p}S_p \\ X_2 &= a_{21}S_1 + a_{22}S_2 + \cdots + a_{2p}S_p \\ &\vdots \qquad\qquad\qquad \vdots \\ X_p &= a_{p1}S_1 + a_{p2}S_2 + \cdots + a_{pp}S_p \end{aligned} \tag{3}$$

However, for any orthogonal matrix $\boldsymbol{R}$, we can write

$$\begin{aligned} X &= \boldsymbol{A}S \\ &= \boldsymbol{A}\boldsymbol{R}^T\boldsymbol{R}S \\ &= \boldsymbol{A}^*S^* \end{aligned} \tag{4}$$

Hence there are many such decompositions and it is therefore impossible to identify any particular latent variable as unique underlying sources. The classical factor analysis model has the form $(q < p)$

$$
\begin{aligned}
X_1 &= a_{11}S_1 + a_{12}S_2 + \cdots + a_{1q}S_q + \varepsilon_1 \\
X_2 &= a_{21}S_1 + a_{22}S_2 + \cdots + a_{2q}S_q + \varepsilon_2 \\
&\vdots \qquad\qquad\qquad \vdots \\
X_p &= a_{p1}S_1 + a_{p2}S_2 + \cdots + a_{pq}S_q + \varepsilon_p
\end{aligned}
\tag{5}
$$

or

$$
X = \boldsymbol{A}S + \varepsilon \tag{6}
$$

Typically the S_j and ε_j are modeled as Gaussian random variables and the model is fit by maximum likelihood.

2 Independent Component Analysis

The ICA model has the form:

$$
\begin{aligned}
X_1 &= a_{11}S_1 + a_{12}S_2 + \cdots + a_{1p}S_p \\
X_2 &= a_{21}S_1 + a_{22}S_2 + \cdots + a_{2p}S_p \\
&\vdots \qquad\qquad\qquad \vdots \\
X_p &= a_{p1}S_1 + a_{p2}S_2 + \cdots + a_{pp}S_p
\end{aligned}
\tag{7}
$$

or

$$
X = \boldsymbol{A}S \tag{8}
$$

where the S_i are assumed to be statistically independent rather than uncorrelated. Intuitively, uncorrelated determines the second cross moment Cov(X), and statistically independent determines all cross moments.
We wish to recover the matrix $\boldsymbol{A}$ in $X = \boldsymbol{A}S$. Without loss of generality, we can assume that X has already been whitened to have $\mathrm{Cov}(X) = \mathbf{I}$, which implies that $\boldsymbol{A}$ is orthogonal. Solving the ICA problem amounts to findg and orthogonal $\boldsymbol{A}$ such that the components of the vector random variables $S = \boldsymbol{A}^T X$ are independent and non-Gaussian.
Many of the popular approaches to ICA are based on entropy. The differential entropy H of a random variable Y with density $g(y)$ is given by

$$
H(Y) = E[-\log g(Y)] = -\int g(y) \log g(y) dy \tag{9}
$$

The quantity $I(Y)$ is called the *Kullback-Leibler* distance or *mutual information*

$$
I(Y) = \sum_{j=1}^{p} H(Y_j) - H(Y) \tag{10}
$$

This is the measurement of *Kullback-Leibler* distance betweenthe density $g(y)$ of Y and its independence version $\prod_{j=1}^{p} g_j(y_j)$, where $g_j(y_j)$ is the marginal density of Y_j. Now, if X has covariance $\mathbf{I}$, and $Y = \boldsymbol{A}^T X$ with $\boldsymbol{A}$ orthogonal, then we will have

$$I(Y) = \sum_{j=1}^{p} H(Y_j) - H(Y) \tag{11}$$

$$= \sum_{j=1}^{p} H(Y_j) - H(X) - \log|\det \boldsymbol{A}| \tag{12}$$

$$= \sum_{j=1}^{p} H(Y_j) - H(X) \tag{13}$$

since we have $Y \sim f_Y(y)$, and $Y = \boldsymbol{A}^T X$, where we can get $X = \boldsymbol{A}Y$, and the PDF of X is $f_X(x) = f_Y(x)||\boldsymbol{A}||$

$$H(X) = E[-\log f_X(x)] = -E\left[\log\big(||\boldsymbol{A}|| \cdot f_Y(x)\big)\right] \tag{14}$$

$$= -E\left[\log||A|| + \log f_Y(x)\right] \tag{15}$$

$$= -\log||A|| - E[-\log f_Y(x)] \tag{16}$$

$$= -\log||A|| + H[Y] \tag{17}$$

$$= H[Y] \tag{18}$$