Tutorial on Factor Analysis

Javier R. Movellan

1 Generative Model

$$O = aH + Z \tag{1}$$

where O is an n-dimensional random vector of observations, H and d-dimensional vector of hidden variables and Z a noise vector. H is a standard Gaussian vector, i.e, $H \sim N(0, I_d)$ and Z is a zero-mean Gaussian vector with diagonal covariance matrix Ψ. The model parameters are the mixing matrix a and the covariance matrix Ψ.

2 EM Learning

Given $o^{(1)}, cdots, o^{(s)}$, a fair sample of observations from O, our goal is to values of $\lambda = \{a, \Psi\}$ that maximize $\sum \log p(o^{(i)} \mid \lambda)$. To do so within the EM framework[1] we form an auxiliary function

$$Q(\bar{\lambda}, \lambda) \stackrel{\text{def}}{=} \sum_i \int p(h \mid o^{(i)} \bar{\lambda}) \log p(o^{(i)} h \mid \lambda) dh \tag{2}$$

we note that

$$p(o^{(i)} h \mid \lambda) = p(h) p(o^{(i)} \mid h\lambda) \tag{3}$$

and since $p(h)$ does not depend on λ we redifine Q as follows

$$Q(\bar{\lambda}, \lambda) \stackrel{\text{def}}{=} \sum_i \int p(h \mid o^{(i)} \bar{\lambda}) \log p(o^{(i)} \mid h\lambda) dh \tag{4}$$

From the definition of the multivariate Gaussian distribution it follows that

$$Q(\bar{\lambda}, \lambda) = -\frac{1}{2} \sum_{i=1} E\left[\log\left(2\pi \left|\Psi\right|\right) + (o^{(i)} - aH)^T \Psi^{-1} (o^{(i)} - aH) \middle| o^{(i)} \bar{\lambda} \right] \tag{5}$$

Taking the gradient with respect to a and setting it to zero[2] we get

$$\nabla_a Q(\bar{\lambda}, \lambda) = \Psi^{-1} \sum_{i=1}^{s} E\left[(o^{(i)} - aH) H^T \mid o^{(i)} \bar{\lambda} \right] = 0 \tag{6}$$

Thus

$$\sum_{i=1}^{s} o^{(i)} \left(h^{(i)} \right)^T = a \sum_{i=1}^{s} E\left[HH^T \mid o^{(i)} \bar{\lambda} \right] \tag{7}$$

where

$$h^{(i)} = E\left[H \mid o^{(i)} \bar{\lambda} \right] = E(H \mid \bar{\lambda}) + \Sigma_{HO} \Sigma_{OO}^{-1} (o^{(i)} - E(O \mid \bar{\lambda})) = a^T (aa^T + \Psi)^{-1} o^{(i)} \tag{8}$$

$$E\left[HH^T \mid o^{(i)} \bar{\lambda} \right] = \Sigma_{HH} - \Sigma_{H0} \Sigma_{00}^{-1} \Sigma_{OH} = I_d - a^T (aa^T + \Psi)^{-1} a + h^{(i)} (h^{(i)})^T \tag{9}$$

Thus,

$$\hat{a} = \left(\sum_{i=1}^{s} o^{(i)} b^T (o^{(i)})^T \right) \left(sI_d - sba + \sum_{i=1}^{s} h^{(i)} (h^{(i)})^T \right)^{-1} \tag{10}$$

where

$$b = a^T (aa^T + \Psi)^{-1} \tag{11}$$

---

[1] See the tutorial on EM from the Kolmogorov project

[2] See the tutorial on Matrix Calculus from the Kolmogorov project

which is independent of Ψ. Note this solves the linear regression problem for least-squares prediction of $o^{(i)}$ based on $h^{(i)}$. Using the optimal value of a and gradient with respect to Ψ we get

$$\nabla_{\Psi^{-1}} Q(\bar{\lambda}, \hat{a}, \Psi) = \text{diag}\nabla_{\Psi^{-1}} Q(\bar{\lambda}, \hat{a}, \Psi)$$
$$= \text{diag}\left[-\frac{s}{2}\Psi - \frac{1}{2}\sum_{i=1}^{s} E\left[(o^{(i)} - \hat{a}H)^T(o^{(i)} - \hat{a}H) \;\Big|\; o^i\bar{\lambda}\right]\right] = 0 \tag{12}$$

Thus,

$$\hat{\Psi} = \frac{1}{s}\sum_{i-1}^{s} \text{diag}\left[o^{(i)}(o^{(i)})^T - 2o^{(i)}\left(\hat{a}E\left[H^T \mid o^{(i)}\bar{\lambda}\right]\right)^T + \hat{a}E\left[HH^T \mid o^{(i)}\bar{\lambda}\right]\hat{a}^T\right] \tag{13}$$

3 Independent Factor Analysis

In ICA one uses superGaussian source distributions, as opposed to Gaussians. A convinient way to get a super-Gasussian sources is to use a mixture of 2 Gaussians with zero mean and variances $1, \alpha$. We will define a new random vector $M = (M_1, \cdots, M_d)^T$ where M_i takes values in $\{0, 1\}$. The distribution of S given $M = m$ is zero mean Gaussian with variance

$$\sigma_m = \alpha^2 \text{ diag}(m) + (I_m - \text{diag}(m)) \tag{14}$$

Here we address the case in which $\lambda = \{a, \Psi\}$. To use the EM algorithm we now need to treat H, M as hidden variables. We have

$$p(ohm \mid \lambda) = p(m)p(h \mid m)p(o \mid hm\lambda) \tag{15}$$

Thus,

$$Q(\bar{\lambda}, \lambda) \stackrel{\text{def}}{=} \sum_i \sum_d \int p(hm \,|\, o^{(i)}\bar{\lambda}) \left(\log p(m) + \log p(h \mid m) + \log p(o^{(i)} \mid hm\lambda)\right) dhdm \tag{16}$$

We can redefine Q by eliminating the terms that do not depend on λ. Thus

$$Q(\bar{\lambda}, \lambda) \stackrel{\text{def}}{=} \sum_m \sum_i p(m \mid o^{(i)}\bar{\lambda}) \int p(h \mid o^{(i)}m\bar{\lambda}) \log p(o^{(i)} \mid hm\lambda) dh \tag{17}$$

$$= \sum_m p(m \mid o^{(i)}\bar{\lambda}) Q_m(\bar{\lambda}, \lambda) \tag{18}$$

Taking derivatives with respect to a we get

$$\nabla_a Q(\bar{\lambda}, \lambda) = \Psi^{-1} \sum_m \bar{p}(m \mid o^{(i)}\bar{\lambda}) \sum_{i=1}^{s} E\left[(o^{(i)} - aH)H^T \;\Big|\; o^{(i)}\bar{\lambda}\right] = 0 \tag{19}$$

Thus

$$\sum_m p(m \,|\, o^{(i)}\bar{\lambda}) \sum_{i=1}^{s} o^{(i)}\left(E\left[H \mid o^{(i)}m\bar{\lambda}\right]\right)^T = a\sum_m p(m \,|\, o^{(i)}\bar{\lambda}) \sum_{i=1}^{s} E\left[HH^T \mid o^{(i)}m\bar{\lambda}\right] \tag{20}$$

where

$$E\left[H \mid\ o^{(i)} m\bar{\lambda}\right] = E(H \mid m\bar{\lambda}) + \Sigma^m_{HO}\left(\Sigma^m_{OO}\right)^{-1}\left(o^{(i)} - E(O \mid m\bar{\lambda})\right) = \sigma_m a^T (a\sigma_m a^T + \Psi)^{-1} o^{(i)} \tag{21}$$

$$E\left[HH^T \mid\ o^{(i)} m\bar{\lambda}\right] = \Sigma^m_{HH} - \Sigma^m_{H0}\left(\Sigma^m_{00}\right)^{-1}\Sigma^m_{OH} + E\left[H \mid o^{(i)} m\bar{\lambda}\right]\left(E\left[H \mid o^{(i)} m\bar{\lambda}\right]\right)^T$$

$$= \sigma_m - \sigma_m a^T (a\sigma_m a^T + \Psi)^{-1} a\sigma_m + E\left[H \mid o^{(i)} m\bar{\lambda}\right]\left(E\left[H \mid o^{(i)} m\bar{\lambda}\right]\right)^T \tag{22}$$

and

$$p(m \mid o^{(i)}\bar{\lambda}) = \frac{p(m)p(o^{(i)} \mid m\lambda)}{\sum_{m'} p(m')p(o^{(i)} \mid m\bar{\lambda})} \tag{23}$$

4 History

- The first version of this document was written by Javier R. Movellan in January 2004, as part of the Kolmogorov project.