EM algorithm — Maximum likelihood estimation of the non-standard t-distribution

Lucas Støjko Andersen

October 26, 2022

1 Stochastic Gradient Descent

We try to minimize the loss function

$$\ell(\theta) = \frac{1}{N}\sum_{i=1}^{N}(y_i - f(x_i|\alpha, \beta, \gamma, \rho))^2 \tag{1}$$

with

$$f(x|\alpha, \beta, \gamma, \rho) = \gamma + \frac{\rho - \gamma}{1 + e^{\beta \log x - \alpha}}. \tag{2}$$

1.1 The Gradient

The loss function has the gradient

$$\nabla \ell(\theta) = -\frac{1}{2N}\sum_{i=1}^{N}(y_i - f(x_i|\alpha, \beta, \gamma, \rho))\nabla f(x_i|\alpha, \beta, \gamma, \rho).$$

We have

$$\begin{aligned}
\nabla_\alpha f &= \frac{\rho - \gamma}{(1 + e^{\beta \log x - \alpha})^2} e^{\beta \log x - \alpha} \\
\nabla_\beta f &= -\frac{\rho - \gamma}{(1 + e^{\beta \log x - \alpha})^2} e^{\beta \log x - \alpha} \log x \\
\nabla_\gamma f &= 1 - \frac{1}{1 + e^{\beta \log x - \alpha}} \\
\nabla_\rho f &= \frac{1}{1 + e^{\beta \log x - \alpha}}.
\end{aligned}$$

1.2 The Hessian

We have the hessian of f

$$
\begin{aligned}
\nabla_{\gamma^2} f &= 0 = \nabla_{\rho^2} f = \nabla_{\rho\gamma} f \\
\nabla_{\alpha^2} f &= 2\frac{\rho-\gamma}{(1+e^{\beta\log x-\alpha})^3}\left(e^{\beta\log x-\alpha}\right)^2 - \frac{\rho-\gamma}{(1+e^{\beta\log x-\alpha})^2}e^{\beta\log x-\alpha} \\
\nabla_{\beta^2} f &= 2\frac{\rho-\gamma}{(1+e^{\beta\log x-\alpha})^3}\left(\log x\; e^{\beta\log x-\alpha}\right)^2 - \frac{\rho-\gamma}{(1+e^{\beta\log x-\alpha})^2}(\log x)^2 e^{\beta\log x-\alpha} \\
\nabla_{\alpha\beta} f &= -\log x\; \nabla_{\alpha^2} f \\
\nabla_{\alpha\rho} f &= \frac{1}{(1+e^{\beta\log x-\alpha})^2}e^{\beta\log x-\alpha} \\
\nabla_{\alpha\gamma} f &= 1 - \nabla_{\alpha\rho} f \\
\nabla_{\beta\rho} f &= -\frac{1}{(1+e^{\beta\log x-\alpha})^2}\log x\; e^{\beta\log x-\alpha} \\
\nabla_{\beta\gamma} f &= 1 - \nabla_{\beta\rho} f
\end{aligned}
$$