Department of Mathematics
Technical University of Munich
Elisabeth Ullmann

October 10, 2024

Solution Sheet 2 – Cross-entropy method

Autumn School *Uncertainty Quantification for High-Dimensional Problems*

Problem 1 (Cross-entropy method with Gaussian densities)

Consider the cross-entropy (CE) method with a family of n-variate Gaussian densities, that are parameterized by their mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and invertible covariance matrix $\Sigma \in \mathbb{R}^{n\times n}_{sym}$. That is, the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathrm{vec}(\Sigma))^\top$ and

$$g(\boldsymbol{u};\boldsymbol{\theta}) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}(\boldsymbol{u}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{u}-\boldsymbol{\mu})\right), \quad \boldsymbol{u} \in \mathbb{R}^n.$$

Let $\boldsymbol{u}_j^{(i)}$, $i = 1, \ldots, N$ denote samples of $g(\cdot;\boldsymbol{\theta}_j)$. Following the lecture notes we define the objective function

$$J(\boldsymbol{\theta};\gamma_j,\boldsymbol{\theta}_j) := \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{G\le\gamma_j\}}(\boldsymbol{u}_j^{(i)}) W(\boldsymbol{u}_j^{(i)};\boldsymbol{\theta}_0,\boldsymbol{\theta}_j) \ln g(\boldsymbol{u}_j^{(i)};\boldsymbol{\theta}),$$

where $W(\boldsymbol{u};\boldsymbol{\theta}_0,\boldsymbol{\theta}_j) = g(\boldsymbol{u};\boldsymbol{\theta}_0)/g(\boldsymbol{u};\boldsymbol{\theta}_j)$. Let

$$H_j^{(i)} := \mathbb{1}_{\{G\le\gamma_j\}}(\boldsymbol{u}_j^{(i)}) W(\boldsymbol{u}_j^{(i)};\boldsymbol{\theta}_0,\boldsymbol{\theta}_j), \quad i = 1, \ldots, N.$$

Show that the solution of the optimization problem

$$J(\boldsymbol{\theta};\gamma_j,\boldsymbol{\theta}_j) \to \max_{\boldsymbol{\theta}}!$$

is given by

$$\begin{aligned}
\boldsymbol{\mu}_{opt} &= \frac{\sum_{i=1}^N H_j^{(i)} \boldsymbol{u}_j^{(i)}}{\sum_{i=1}^N H_j^{(i)}}, \\
\Sigma_{opt} &= \frac{\sum_{i=1}^N H_j^{(i)} (\boldsymbol{u}_j^{(i)} - \boldsymbol{\mu}_{opt})(\boldsymbol{u}_j^{(i)} - \boldsymbol{\mu}_{opt})^\top}{\sum_{i=1}^N H_j^{(i)}}.
\end{aligned}$$

Hint. It holds
$$\frac{\partial \det(\Sigma)}{\partial \Sigma} = \det(\Sigma)\Sigma^{-\top}.$$

Solution

Recall that $\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathrm{vec}(\Sigma))^\top$. Moreover, it holds

$$\ln g(\boldsymbol{u}; \boldsymbol{\theta}) = -\frac{1}{2}(\boldsymbol{u} - \boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{u} - \boldsymbol{\mu}) - \frac{1}{2}\ln(\det(2\pi\Sigma)).$$

Step 1. The gradient of $\ln g(\boldsymbol{u}; \boldsymbol{\theta})$ with respect to $\boldsymbol{\mu}$ is given by

$$\nabla_{\boldsymbol{\mu}} \ln g(\boldsymbol{u}; \boldsymbol{\theta}) = \Sigma^{-1}(\boldsymbol{u} - \boldsymbol{\mu}).$$

The necessary optimality condition $\nabla_{\boldsymbol{\mu}} J(\boldsymbol{\theta}; \gamma_j, \boldsymbol{\theta}_j) = \mathbf{0}$ gives

$$\sum_{i=1}^N H_j^{(i)} \Sigma_{opt}^{-1}(\boldsymbol{u}_j^{(i)} - \boldsymbol{\mu}_{opt}) = \mathbf{0}.$$

Since Σ_{opt} is independent of i, we can multiply both sides of the above equation from the left by Σ_{opt} and obtain

$$\sum_{i=1}^N H_j^{(i)} (\boldsymbol{u}_j^{(i)} - \boldsymbol{\mu}_{opt}) = \mathbf{0}.$$

This shows the claim

$$\boldsymbol{\mu}_{opt} = \frac{\sum_{i=1}^N H_j^{(i)} \boldsymbol{u}_j^{(i)}}{\sum_{i=1}^N H_j^{(i)}}.$$

Step 2. The gradient of $\ln g(\boldsymbol{u}; \boldsymbol{\theta})$ with respect to Σ is given by

$$\nabla_\Sigma \ln g(\boldsymbol{u}; \boldsymbol{\theta}) = -\frac{1}{2}\left(\Sigma^{-1} - \Sigma^{-1}(\boldsymbol{u} - \boldsymbol{\mu})(\boldsymbol{u} - \boldsymbol{\mu})^\top \Sigma^{-1}\right). \tag{1}$$

Consider an invertible matrix $\boldsymbol{Y} = \boldsymbol{Y}(x)$, then it holds

$$\frac{\partial \boldsymbol{Y}^{-1}}{\partial x} = -\boldsymbol{Y}^{-1}\frac{\partial \boldsymbol{Y}}{\partial x}\boldsymbol{Y}^{-1}.$$

From this it follows

$$\frac{\partial \Sigma_{k\ell}^{-1}}{\partial \Sigma_{ij}} = -\Sigma_{ki}^{-1}\Sigma_{j\ell}^{-1}$$

and

$$\frac{\partial \boldsymbol{v}^\top \Sigma^{-1} \boldsymbol{v}}{\partial \Sigma} = -\Sigma^{-\top}\boldsymbol{v}\boldsymbol{v}^\top\Sigma^{-\top}.$$

Using the hint we obtain

$$\frac{\ln(\det(2\pi\Sigma))}{\partial \Sigma} = \frac{1}{\det(2\pi\Sigma)}\det(2\pi\Sigma)(2\pi\Sigma)^{-\top}2\pi = \Sigma^{-\top}.$$

This shows (1). The necessary optimality condition $\nabla_\Sigma J(\boldsymbol{\theta}; \gamma_j, \boldsymbol{\theta}_j) = 0$ gives

$$\sum_{i=1}^N H_j^{(i)}\left(\Sigma_{opt}^{-1} - \Sigma_{opt}^{-1}(\boldsymbol{u}_j^{(i)} - \boldsymbol{\mu}_{opt})(\boldsymbol{u}_j^{(i)} - \boldsymbol{\mu}_{opt})^\top \Sigma_{opt}^{-1}\right) = \mathbf{0}.$$

Since Σ_{opt} is independent of i, we can multiply both sides of the above equation from the left and from the right by Σ_{opt} and obtain

$$\sum_{i=1}^{N} H_j^{(i)} \left(\Sigma_{opt} - (\boldsymbol{u}_j^{(i)} - \boldsymbol{\mu}_{opt})(\boldsymbol{u}_j^{(i)} - \boldsymbol{\mu}_{opt})^\top \right) = \mathbf{0}.$$

This shows the claim

$$\Sigma_{opt} = \frac{\sum_{i=1}^{N} H_j^{(i)} (\boldsymbol{u}_j^{(i)} - \boldsymbol{\mu}_{opt})(\boldsymbol{u}_j^{(i)} - \boldsymbol{\mu}_{opt})^\top}{\sum_{i=1}^{N} H_j^{(i)}}.$$

Problem 2 (Implementation of CE method)

For $\boldsymbol{u} \in \mathbb{R}^2$ consider the limit-state function $G(\boldsymbol{u}) = -\min(u_1, u_2) + a$, where $a > 0$. Suppose that the nominal density f is the bivariate standard normal density. In this case $P_f = \Phi(-a)^2$.

The Matlab file `CE_Gaussian_family.m` implements the CE method with a family of Gaussian densities parameterized by their mean value $\boldsymbol{\mu} \in \mathbb{R}^2$ and covariance matrix $\Sigma \in \mathbb{R}^{2\times 2}$.

a) In which lines do we use the analytical updates for $\boldsymbol{\mu}_{opt}$ and Σ_{opt} derived in Problem 1?

b) Compare the estimate for P_f obtained with standard Monte Carlo and the CE method for different values of a! What do you conclude?

Solution

See the file `CE_Gaussian_family.m`.