Cross Entropy Method

Tomasz Lamża

24 kwietnia 2022

1 CE algorithm for optimization

1. Choose an initial parameter vector $\hat{\mathbf{v}}_0$. Let $N^e = \lceil \varrho N \rceil$. Set t = 1 (level counter). Where ϱ is a user specified parameter called the rarity parameter.

2. Generate $\mathbf{X}_1, ..., \mathbf{X}_N \sim_{iid} f(\cdot; \hat{\mathbf{v}}_{t-1})$. Calculate the performances $S(\mathbf{X_i})$ for all i, and order them from smallest to largest: $S_{(1)} \leqslant ... \leqslant S_{(N)}$. Let $\hat{\gamma}_t$ be sample (1-ϱ) - quantile of performances; that is, $\hat{\gamma}_t = S_{(N-N^e+1)}$.

3. Use the **same** sample $\mathbf{X}_1, ..., \mathbf{X}_N$ and solve the stochastic program

$$\max_{\mathbf{v}} \hat{D}(\mathbf{v}) = \max_{\mathbf{v}} \frac{1}{N} \sum_{k=1}^{N} I_{\{S(\mathbf{X}) \geqslant \hat{\gamma}_t\}} \ln f(\mathbf{X}_k; \mathbf{v}). \tag{1}$$

   Denote the solution by $\hat{\mathbf{v}}_t$. Symbol I represents indicator function - it maps elements of the subset to one, and all other elements to zero. In case above if $S(\mathbf{X})$ is grater or equal $\hat{\gamma}_t$ then I will return 1, else it will return 0. The In most cases the function $\hat{D}$ is convex and differentiable with respect to $\mathbf{v}$ and, thus, the solution may be readily obtain by solving (with respect to $\mathbf{v}$) the following system of equations:

$$\frac{1}{N} \sum_{k=1}^{N} I_{\{S(\mathbf{X}) \geqslant \hat{\gamma}_t\}} \bigtriangledown \ln f(\mathbf{X}_k; \mathbf{v}) = 0 \tag{2}$$

   The solution of (2) can often be calculated analytically. In particular, this happens if the distributions of the random variables belong to a natural exponential family (NEF).

4. If some stopping criterion is met, stop; otherwise, set $t = t + 1$, and return to Step 2.

Example of (2) with normal distribution

$$f(\mathbf{x}; \sigma, \mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

$$\ln f(\mathbf{x}; \sigma, \mu) = -\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2 + \ln \frac{1}{\sigma\sqrt{2\pi}}$$

$$\frac{\partial \ln f(\mathbf{x};\sigma,\mu)}{\partial \sigma} = \frac{\partial \ln \frac{1}{\sigma\sqrt{2\pi}}}{\partial \sigma} - \frac{\partial \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}{\partial \sigma} = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3}$$

$$\frac{\partial \ln f(\mathbf{x};\sigma,\mu)}{\partial \mu} = \frac{\mu - x}{\sigma^2}$$

And now (2) can be calculated:

$$(2) = \frac{1}{N}\sum_{k=1}^{N} I_{\{S(\mathbf{X}_k)\geqslant\hat{\gamma}_t\}} \begin{bmatrix} \frac{\partial \ln f(\mathbf{X}_k;\sigma,\mu)}{\partial \mu} \\ \frac{\partial \ln f(\mathbf{X}_k;\sigma,\mu)}{\partial \sigma} \end{bmatrix} = \frac{1}{N}\sum_{k=1}^{N} I_{\{S(\mathbf{X}_k)\geqslant\hat{\gamma}_t\}} \begin{bmatrix} -\frac{1}{\sigma} + \frac{(\mu-\mathbf{X}_k)^2}{\sigma^3} \\ \frac{\mu-\mathbf{X}_k}{\sigma^2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad (3)$$

Let $\hat{N}$ be the set of indexes of elements that meet the criteria $S(\mathbf{X}_k) \geqslant \hat{\gamma}_t$. Then

$$\frac{1}{N}\sum_{k\in\hat{N}} \begin{bmatrix} -\frac{1}{\sigma} + \frac{(\mu-\mathbf{X}_k)^2}{\sigma^3} \\ \frac{\mu-\mathbf{X}_k}{\sigma^2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\sum_{k\in\hat{N}} \begin{bmatrix} -1 + \frac{(\mu-\mathbf{X}_k)^2}{\sigma^2} \\ \mu - \mathbf{X}_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} \sum_{k\in\hat{N}}(-1 + \frac{(\mu-\mathbf{X}_k)^2}{\sigma^2}) & = 0 \\ \sum_{k\in\hat{N}}(\mu - \mathbf{X}_k) & = 0 \end{cases}$$

$$\begin{cases} |\hat{N}|\sigma^2 & = \sum_{k\in\hat{N}}(\mu - \mathbf{X}_k)^2 \\ \mu & = \mathrm{avg}(\sum_{k\in\hat{N}} \mathbf{X}_k) \end{cases}$$

$$\begin{cases} \sigma & = \sqrt{\frac{\sum_{k\in\hat{N}}(\mathbf{X}_k-\mu)^2}{|\hat{N}|}} \\ \mu & = \mathrm{avg}(\sum_{k\in\hat{N}} \mathbf{X}_k) \end{cases}$$