Discussion regarding the optimization task based on constrained minimum cross entropy

Vladimíra Sečkárová

Seminar AS

08.10.2011

Optimization task

We are looking for the solution of the following optimization task (constrained minimum cross entropy):

$$\min_{\pi} \int_{H\times\ldots\times H} \pi(h_{1:n}|D) \log \frac{\pi(h_{1:n}|D)}{q(h_{1:n})} dh(x_1)\ldots dh(x_n) \qquad (1)$$

s.t.

$$\mathrm{E}_{\pi(h_{1:n}|D)}[\mathrm{KL}(g_j|h_{1:n})|D] \leq \beta_j \quad j = 1,\ldots,s, \qquad (2)$$

where KL stands for the Kullback-Leibler divergence and $h_{1:n} = (h(x_1),\ldots,h(x_n))$.

Solution

Under the following choice of the prior distribution $q(h_{1:n})$ (Dirichlet $Dir(1/n, \ldots, 1/n)$):

$$q(h_{1:n}) = \frac{1}{\frac{(\Gamma(1/n))^n}{\Gamma(1)}} \prod_{i=1}^{n} h(x_i)^{1/n-1}$$

the solution of the optimization task (1) s.t. (2) is:

$$\hat{\pi}(h_{1:n}|D) = \frac{1}{\frac{\prod_{i=1}^{n} \Gamma(1/n+\sum_j \lambda_j g_j(x_i))}{\Gamma(1+\sum_j \lambda_j)}} \prod_{i=1}^{n} h(x_i)^{1/n+\sum_j \lambda_j g_j(x_i)-1} \qquad (3)$$

which stands for the probability density function (pdf) of the Dirichlet distribution $Dir(1/n + \sum_j \lambda_j g_j(x_1), \ldots)$. Clearly, the parameters of determined distribution depend on Lagrange multipliers $\lambda_1, \ldots, \lambda_n$. Next step consists of their computation.

Insert computed distribution into constraints

1. Insert the determined pdf (3) into the constraints (2), for j^{th} constraint it looks as follows:

$$\mathrm{E}_{\hat{\pi}(h_{1:n}|D)}[\mathrm{KL}(g_j|h_{1:n})|D] = \tag{4}$$

$$\int_H \frac{\Gamma(1+\sum_j \lambda_j)}{\prod_{i=1}^n \Gamma(1/n + \sum_j \lambda_j g_j(x_i))} \prod_{i=1}^n h(x_i)^{1/n+\sum_j \lambda_j g_j(x_i)-1}$$

$$\times \left(\sum_{l=1}^n g_j(x_l) \log \frac{g_j(x_l)}{h(x_l)} \right) dh(x_1) \ldots dh(x_n)$$

Compute the expectations

2. After reformulation we get that (4) equals (for a particular constraint):

$$-H(g_j)+\sum_{l=1}^{n} g_j(x_l)\left(\psi(1+\sum_{j}\lambda_j)-\psi(1/n+\sum_{j}\lambda_j g_j(x_l))\right)$$

where $H(.)$ stands for the entropy.

Compute λ's

- For computation of Lagr. multipliers $\lambda_1, \ldots$ we use Nelder Mead method.
- Aim: minimize the objective function in the multidimensional space.
- Nelder Mead method: approximates the local optimum by extrapolating the behavior of the objective function by using the concept of simplex.
- Our objective function:

$$\sum_{j=1}^{s} \mathrm{E}_{\hat{\pi}(h_{1:n}|D)}[\mathrm{KL}(g_j|h_{1:n})|D]$$

- *n*-simplex: n-dimensional polytope which is the convex hull of its n + 1 vertices.

Available tools

- Available tools:
  Nelder Mead program (by Jeff Borggaard): The user set the initial n-simplex and let the method compute the rest.

  Or

  Use the Matlab function where the user set just one initial point, the method computes the rest.

Results

- Nelder Mead with full initial *n*-simplex set by user: results differ in the 3^{rd} decimal place.
  Q: How to set the initial set of points?
- Matlab Nelder Mead (one initial point): result is reasonable.
  Q: Is the initial point $(\lambda_1, \ldots, \lambda_s) = (0, \ldots, 0)$ optional?