Description

1 Problem Setup

Denote the design (input) space as $\mathcal{X}$. Given M models $\{\mathcal{M}_i\}_{i=1}^M$, each with parameters $\theta_i \subseteq \Theta_i$ and prior distribution $p(\mathcal{M}_i)$, we first give each θ_i a (multivariate Gaussian) prior $p(\theta_i|\mathcal{M}_i)$. For simplicity, we assume that one of $\{\mathcal{M}_i\}_{i=1}^M$ is the ground-truth, i.e., $\mathcal{M}_{\text{true}}$.

2 Input Selection Criterion

2.1 Method 1: `getSelCritLogDet.m`

We first draw several samples from the density $p(\theta_i|\mathcal{M}_i)$, denoted by $\{\theta_i^s\}_{s=1}^{K_i}$, using HMC. (As an alternative approach, we find a local minimum of $p(\theta_i|\mathcal{M}_i)$, denoted by θ_i^{MAP}, using HMC.) Then, we estimate the response $y_i^s(x) \triangleq \mathcal{M}_i(x;\theta_i^s) + \epsilon_i^s$, where $\{\epsilon_i^s\}_{i\in[M],s\in[K_i]} \overset{iid}{\sim} \mathcal{N}(0,\sigma_{\text{n}}^2)$. Thus $y_i^s(x) \sim \mathcal{N}(\mathcal{M}_i(x;\theta_i^s),\sigma_{\text{n}}^2)$. For any $(i,s) \in [M]\times[K_i]$ and $(j,t)\in[M]\times[K_j]$, compute

$$\begin{aligned} D_{(i,s),(j,t)}(x) &\triangleq D_{\text{KL}}\left(\mathcal{N}(\mathcal{M}_i(x;\theta_i^s),\sigma_{\text{n}}^2), \mathcal{N}(\mathcal{M}_j(x;\theta_j^t),\sigma_{\text{n}}^2)\right) \\ &= \frac{\Big(\mathcal{M}_i(x;\theta_i^s),\sigma_{\text{n}}^2) - \mathcal{M}_j(x;\theta_j^t),\sigma_{\text{n}}^2)\Big)^2}{2\sigma_{\text{n}}^2}. \end{aligned}$$

We choose the design point x^* to be a local minimum of

$$S(x) \triangleq -\log\det D(x).$$

2.2 Method 2: `getSelCritJSDiv.m`

This method was proposed in [1]. The first step is the same as those in Section 2.1, i.e., we draw several samples from the density $p(\theta_i|\mathcal{M}_i)$, denoted by $\{\theta_i^s\}_{s=1}^{K_i}$, using HMC. For each model $\mathcal{M}_i$, we aim to find the distribution of the its predicted response given x, i.e.,

$$p(y|\mathcal{M}_i,x) = \int_{\Theta_i} p(y|\theta_i,\mathcal{M}_i,x)p(\theta_i|\mathcal{M}_i)\,\mathrm{d}\theta_i.$$

We can approximate this density using the samples $\{\theta_i^s\}_{s=1}^{K_i}$, i.e.,

$$p(y|\mathcal{M}_i,x) \approx \frac{1}{K_i}\sum_{s=1}^{K_i} p(y|\theta_i^s,\mathcal{M}_i,x) = \frac{1}{K_i}\sum_{s=1}^{K_i}\mathcal{N}(\mathcal{M}_i(x;\theta_i^s),\sigma_{\text{n}}^2). \tag{1}$$

The OED criterion is based on the Jensen-Shannon divergence (JSD), i.e.,

$$D_{\mathrm{JS}}(x) \triangleq \mathbb{E}_{\mathcal{M}}\big[D_{\mathrm{KL}}\big(p(y|\mathcal{M},x) \| \mathbb{E}_{\mathcal{M}}[p(y|\mathcal{M},x)]\big)\big] \tag{2}$$

$$= \sum_{i=1}^{n} p(\mathcal{M}_i) D_{\mathrm{KL}}\big(p(y|\mathcal{M}_i,x) \| \textstyle\sum_{i=1}^{n} p(\mathcal{M}_i) p(y|\mathcal{M}_i,x)\big). \tag{3}$$

Then we find a local maximum of $D_{\mathrm{JS}}(x)$ on $\mathcal{X}$, denoted by x^*.

2.3 Method 3: `getSelCritJSDivU.m`

Note that Method 1 in Section 2.1 is ad-hoc and not well-justified. A more principled approach would be as follows. We first approximate $p(y|\mathcal{M}_i,x)$ for each model $\mathcal{M}_i$ as in (1). Then, instead of using the JSD criterion as in (3), we use the weighted sum of pairwise KL divergences of $\{p(y|\mathcal{M}_i,x)\}_{i=1}^{M}$. Specifically, define

$$\widetilde{D}_{\mathrm{KL}}(x) \triangleq \sum_{i,j=1}^{M} p_i p_j D_{\mathrm{KL}}\big(p(y|\mathcal{M}_i,x) \| \, p(y|\mathcal{M}_j,x)\big), \tag{4}$$

and we find a local maximum of $\widetilde{D}_{\mathrm{KL}}(x)$ on $\mathcal{X}$. Note that by Jensen's inequality, $\widetilde{D}_{\mathrm{KL}}(x) \geq D_{\mathrm{JS}}(x)$, for any $x \in \mathcal{X}$.

3 Posteriors of Model and Model Parameters

Then we simulate the response at x^*, i.e., $y(x^*)$, by

$$y(x^*) \triangleq \mathcal{M}_{\mathrm{true}}(x^*) + \epsilon,$$

where $\epsilon \sim \mathcal{N}(0, \sigma_{\mathrm{n}}^2)$. With the data pair $(x^*, y(x^*))$, we obtain the log-likelihood (up to some constants)

$$\log p\left((x^*, y(x^*))|\theta_i, \mathcal{M}_i\right) = -\frac{\left(y(x^*) - \mathcal{M}_i(x^*;\theta_i)\right)^2}{2\sigma_{\mathrm{n}}^2}.$$

Then the log-posterior (again, up to some constants) is given by

$$\log p\left(\theta_i|(x^*, y(x^*)), \mathcal{M}_i\right) = \log p(\theta_i|\mathcal{M}_i) + \log p\left((x^*, y(x^*))|\theta_i, \mathcal{M}_i\right).$$

In addition, the likelihood of the model $\mathcal{M}_i$ is

$$p\left((x^*, y(x^*))|\mathcal{M}_i\right) = \int_{\Theta_i} p\left((x^*, y(x^*))|\theta_i, \mathcal{M}_i\right) p(\theta_i|\mathcal{M}_i) \, \mathrm{d}\theta_i,$$

which again, can be evaluated by HMC. Then we obtain the posterior distribution over $\{\mathcal{M}_i\}_{i=1}^{M}$ as

$$p\left(\mathcal{M}_i|(x^*, y(x^*))\right) = \frac{p\left((x^*, y(x^*))|\mathcal{M}_i\right) p(\mathcal{M}_i)}{\sum_{i=1}^{M} p\left((x^*, y(x^*))|\mathcal{M}_i\right) p(\mathcal{M}_i)}.$$

4 Test Model

We take equation (I.24.6) from Feynman's lecture notes, which is

$$E = cm^{e_1}(\omega^{e_2} + \omega_0^{e_3})z^{e_4}, \tag{5}$$

where $c = 1/4$, $e_1 = 1$ and $e_2 = e_3 = e_4 = 2$. This model has four inputs $x \triangleq (m, \omega, \omega_0, z)$ and five parameters $\theta \triangleq (c, e_1, e_2, e_3, e_4)$. We use three candidate models, the first of which is the ground-truth model in (5). The other two models are

$$E = cm^{e_1}\omega^{e_2}\omega_0^{e_3}z^{e_4}, \tag{6}$$

$$E = cm^{e_1}(\omega^{e_2} + z^{e_4})\omega_0^{e_3}. \tag{7}$$

We can encode the initial values of the parameters of each model, say θ_i in $\mathcal{M}_i$, in the prior distribution $p(\theta_i|\mathcal{M}_i)$.

References

[1] J. Vanlier, C. A. Tiemann, P. A. Hilbers, and N. A. van Riel, "Optimal experiment design for model selection in biochemical networks," *BMC Syst. Biol.*, vol. 8, no. 1, 2014.