BOED

Bayesian optimal experimental design

parameter θ design ξ (x) prior $p(\theta)$

θ, ξ $\rightarrow$ y (outcome)

joint $p(y,\theta|\xi) = \underbrace{p(y|\theta,\xi)}_{\text{likelihood}}\, p(\theta)$ θ not depend on ξ

explicit / implicit $= p(y|\xi)\, p(\theta|\xi,y)$ $\rightarrow$ posterior

$$p(y|\xi) \rightarrow \mathbb{E}_{p(\theta)}[p(y|\xi,\theta)]$$

double intractable: both posterior and its entropy are hard to compute

Info Gain $IG(\xi,y) = \mathcal{H}[p(\theta)] - \mathcal{H}[p(\theta|\xi,y)]$

Expected IG (EIG) $I(\xi) = \mathbb{E}_{p(y|\xi)}[IG(\xi,y)] = MI(\theta;y|\xi)$

$$= \mathbb{E}_{p(y|\xi)p(\theta|\xi,y)}[\log p(\theta|\xi,y) - \log p(\theta)] = -\mathbb{E}_{p(y|\xi)}[\mathcal{H}[p(\theta|\xi,y)]] + \mathcal{H}[p(\theta)]$$

$$= \mathbb{E}_{p(\theta)p(y|\theta,\xi)}[\log p(y|\theta,\xi) - \log p(y|\xi)] = -\mathbb{E}_{p(\theta)}[\mathcal{H}[p(y|\theta,\xi)]] + \mathcal{H}[p(y|\xi)]$$

optimal design $\xi^* = \arg\max_\xi I(\xi)$

Explicit likelihood

estimate by IS

$$\hat{p}(y|\xi) \approx \frac{1}{N}\sum_{n=1}^N p(y|\theta_n,\xi) \qquad \theta_n \overset{iid}{\sim} p(\theta)$$

$$\mathbb{E}_{p(\theta|\xi,y)}[\log p(y|\theta,\xi)] \approx \frac{1}{N}\sum_{n=1}^N \frac{p(y|\theta_n,\xi)}{\frac{1}{N}\sum_{n'=1}^N p(y|\theta_{n'},\xi)} \log p(y|\theta_n,\xi) \qquad y \sim p(y|\xi)$$

estimate by NMC (nested)

$$I(\xi) \approx \frac{1}{N}\sum_{n=1}^N \log \frac{p(y_n|\theta_n,\xi)}{\hat{p}(y_n|\xi)} \qquad (\theta_n, y_n) \sim p(\theta)p(y|\theta,\xi)$$

$$\approx \frac{1}{N}\sum_{n=1}^N \log \frac{p(y_n|\theta_n,\xi)}{\frac{1}{M}\sum_{m=1}^M p(y_n|\theta'_m,\xi)} \qquad \theta'_m \sim p(\theta)$$

Implicit likelihood

ABC $\theta \sim p(\theta|y,\xi) \iff \tilde{y} \sim p(\cdot|\theta,\xi)$, if $\|y-\tilde{y}\| < \varepsilon$ then accept θ

LFIRE logistic regression $r(\xi,\theta,y) \rightarrow p(y|\theta,\xi)/p(y|\xi)$

$$\Rightarrow I(\xi) = \mathbb{E}_{p(\theta)p(y|\theta,\xi)}[\log r(\xi,\theta,y)] \approx \frac{1}{N}\sum_{n=1}^N \log \hat{r}(\xi,\theta_n,y_n)$$

Bayesian exp design: design ξ $\downarrow$ | observation y $\downarrow$ | param θ $\downarrow$ | design criterion $I(\xi)$ $\downarrow$

SS

Bayesian active learning: query $\downarrow$ | label | | acquisition function α

Pooled-based active learning: unlabeled data | | classifier NN param

$p(y|\theta,\xi)$: Dropout / deep ensemble

Alg: For $t=1,2,\dots$ do:

$$\xi_t = \arg\max_{\xi\in\Xi} \alpha(\xi; D_{t-1}) \quad \text{or} \quad \alpha(\xi; \theta_{t-1})$$

very similar to

$$y_t = \text{Human Labeller}(\xi_t)$$

black box-opt!

$$D_t = D_{t-1} \cup \{(\xi_t, y_t)\}$$

(goal is different)

Calculate $\theta_t \sim p(\theta|D_t)$ or $\theta_t = \arg\max p(\theta|D_t)$

greedy acquisition

$$I(\xi; D_t) = \mathbb{E}_{p(y|\xi,D_t)}\mathbb{E}_{p(\theta|\xi,y,D_t)}[\log p(\theta|\xi,y,D_t)] - \mathbb{E}_{p(\theta|D_t)}[\log p(\theta|D_t)]$$

$$\xi_t = \arg\max_\xi I(\xi; D_t)$$

||

=

Bayesian active learning by disagreement (BALD score) $\equiv$ EIG

$$\alpha_{BALD}(\xi; D_t) = \mathbb{E}_{p(\theta|D_t)}[\mathcal{H}[p(y|\xi,D_t)] - \mathcal{H}[p(y|\xi,\theta,D_t)]]$$

$$= \mathbb{E}_{p(\theta|D_t)p(y|\xi,\theta,D_t)}[\log p(y|\xi,\theta,D_t) - \log p(y|\xi,D_t)] = I(\xi; D_t)$$

Batch BALD

Stepwise Uncertainty Reduction

$$p(\theta|\xi,y,D_t)$$

$$\alpha_{SUR}(\xi; D_t) = -\mathbb{E}_{p(y|\xi,D_t)}[\mathcal{H}[p(\theta|D_t \cup \{(\xi,y)\})]]$$

$$= I(\xi; D_t) + \mathbb{E}_{p(\theta|D_t)}[\log p(\theta|D_t)] = I(\xi; D_t) - \mathcal{H}[p(\theta|D_t)]$$

constant w.r.t. ξ

Entropy Search

$$\alpha_{ES}(\xi; D_t) = \mathbb{E}_{p(y|\xi,D_t)}[KL[p(\theta|D_t \cup \{(\xi,y)\}) \| b(\theta)]]$$

$$= \mathbb{E}_{p(y|\xi,D_t)p(\theta|\xi,y,D_t)}[\log p(\theta|\xi,y,D_t) - \log p(\theta) + \log p(\theta) - \log b(\theta)]$$

$$= I(\xi; D_t) + KL[p(\theta) \| b(\theta)]$$

constant w.r.t. ξ

Other acquistion functions

entropy acquisition $\alpha(\xi; D_t) = \mathcal{H}[p(\underbrace{y|\xi, D_t}_{\mathbb{E}_{p(\theta|D_t)}[p(y|\theta,\xi)]})]$

mean-STD acquisition $\alpha(\xi; D_t) = \frac{1}{|y|}\sum_y \sqrt{Var_{p(\theta|D_t)}[p(y|\xi, D_t)]}$

Prob of improvement $\alpha_{PI}(\xi; D_t) = \mathbb{P}(\overset{GP}{f(\xi)} < \tau | D_t)$ $\tau_t = \max\{y_1, \dots, y_t\}$

Exepcted improvement $\alpha_{EI}(\xi; D_t) = \mathbb{E}[(f(\xi) - \tau)_+ | D_t]$

UCB $\alpha_{UCB\text{-}p}(\xi; D_t) = \overset{\text{p-quantile}}{q_p}(f(\xi) | D_t) = \mu(\xi|D_t) + \overset{\text{p-quantile of } N(0,I)}{\beta_p} \cdot \sigma(\xi|D_t)$

Thompson Sampling $\alpha_{TS}(\xi; D_t) = f_t(\xi)$ $f_t \sim \overset{GP}{p}(f|D_t)$

history $h_{t-1} = \{(\xi_i, y_i)\}_{i=1:t-1}$ $p(y_t|\xi) = \mathbb{E}_{p(\theta|h_{t-1})}[p(y_t|\theta,\xi)]$

$$I_{h_{t-1}}(\xi) = \mathbb{E}_{p(\theta|h_{t-1})p(y_t|\theta,\xi,\cancel{h_{t-1}})}[\log p(y_t|\theta,\xi,\cancel{h_{t-1}}) - \log p(y_t|\xi,h_{t-1})]$$
$$= \quad \cdots \quad [\log p(\theta|h_{t-1},\xi,y_t) - \log p(\theta|h_{t-1})]$$

batch/ static design: determine all $\xi_{1:T}$ before the first iteration $\approx$ one-step w/ larger design space

treating whole sequence as one experiment

variational BOED

BA bound:

$$\hat{\mu}_{posterior}(\xi) = \mathbb{E}_{p(y,\theta|\xi)}[\log q_p(\theta|y,\xi) - \log p(\theta)] \approx \frac{1}{N}\sum_{n=1}^{N}\log\frac{q_p(\theta_n|y_n,\xi)}{p(\theta_n)} \leq I(\xi)$$

$$\hat{\mu}_{marginal}(\xi) = \mathbb{E}_{p(y,\theta|\xi)}[\log p(y|\theta,\xi) - \log q_m(y|\xi)] \approx \frac{1}{N}\sum_{n=1}^{N}\log\frac{p(y_n|\theta_n,\xi)}{q_m(y_n|\xi)} \geq I(\xi)$$

or $q_v(\theta_{1:L}|y)$

$$\hat{\mu}_{VNMC}(\xi) = \mathbb{E}_{p(y,\theta_0|\xi)q_v(\theta_{1:L}|\xi,y)}\left[\log\frac{p(y|\theta_0,\xi)}{\frac{1}{L}\sum_{\ell=1}^{L}\frac{p(y,\theta_\ell|\xi)}{q_v(\theta_\ell|y,\xi)}}\right] \geq I(\xi)$$

IWAE bound, consistent $(L\to\infty)$

$$\Leftarrow \log p(y|\xi) = \log\frac{1}{L}\sum_{\ell=1}^{L}\mathbb{E}_{q_v(\theta_{1:L}|\xi,y)}\left[\frac{p(y,\theta_\ell|\xi)}{q_v(\theta_\ell|y,\xi)}\right] \geq \mathbb{E}_{q_v(\theta_{1:L}|\xi,y)}\left[\log\frac{1}{L}\sum_{\ell=1}^{L}\frac{p(y,\theta_\ell|\xi)}{q_v(\theta_\ell|y,\xi)}\right]$$

$$I_{ACE}(\xi,L) = \mathbb{E}_{p(y,\theta_0|\xi)q_v(\theta_{1:L}|\xi,y)}\left[\log\frac{p(y|\theta_0,\xi)}{\frac{1}{L+1}\sum_{\ell=0}^{L}\frac{p(y,\theta_\ell|\xi)}{q_v(\theta_\ell|y,\xi)}}\right] \leq I(\xi)$$

by $q_v(\theta|y,\xi) \to p(\theta)$

$$I_{PCE}(\xi,L) = \mathbb{E}_{p(y,\theta_0|\xi)p(\theta_{1:L})}\left[\log\frac{p(y|\theta_0,\xi)}{\frac{1}{L+1}\sum_{\ell=0}^{L}p(y|\theta_\ell,\xi)}\right] \leq I(\xi)$$

InfoNCE

(implicit likelihood) $\hat{\mu}_{m+\ell}(\xi) = \mathbb{E}_{p(y,\theta|\xi)}\left[\log\frac{q_\ell(y|\theta,\xi)}{q_m(y|\xi)}\right] \approx \frac{1}{N}\sum_{n=1}^{N}\log\frac{q_\ell(y_n|\theta_n,\xi)}{q_m(y_n|\xi)}$ not a bound of $I(\xi)$

+ sequential

$$p(\theta) \to p(\theta|h_{t-1}) = p(\theta)\prod_{i=1}^{t-1}p(y_i|\theta,\xi_i) \Big/ \prod_{i=1}^{t-1}p(y_i|\xi_i)$$

$$\Rightarrow \hat{\mu}_{post}(\xi_t) = \mathbb{E}_{p(\theta|h_{t-1})p(y_t|\theta,\xi_t)}\left[\log q_p(\theta|y_t,\xi_t) - \log p(\theta)\prod_{i=1}^{t-1}p(y_i|\theta,\xi_i)\right] + \log p(y_{1:t-1}|\xi_{1:t-1})$$

constant for ξ_t

(DAD)

Deep Adaptive Design

design function : $\xi_t = \overset{\text{NN}}{\pi_\phi}(h_{t-1})$

$$MI(\theta; h_T) = \mathcal{H}[p(\theta)] - \mathcal{H}[p(\theta|h_T)] \overset{\text{MI chain rule}}{=} \sum_t \mathbb{E}[\mathcal{H}[p(\theta|h_{t-1})] - \mathcal{H}[p(\theta|h_t)]]$$
$$= \sum_t MI(\theta; (y_t, \xi_t) | h_{t-1})$$

$$\mathcal{I}_T(\pi_\phi) = \mathbb{E}_{p(\theta)p(h_T|\theta,\pi_\phi)}\left[\sum_{t=1}^T I_{h_{t-1}}(\xi_t)\right] \qquad p(h_T|\theta,\pi_\phi) = \prod_t \underbrace{p(y_t|\theta,\xi_t)}_{\text{explicit likelihood}}$$
$$= \mathbb{E}_{p(\theta)p(h_T|\theta,\pi_\phi)}\left[\log \frac{p(h_T|\theta,\pi_\phi)}{p(h_T|\pi_\phi)}\right] \qquad p(h_T|\pi_\phi) = \mathbb{E}_{p(\theta)}[p(h_T|\theta,\pi_\phi)]$$
$$\geq \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi_\phi)p(\theta_{1:L})}\left[\log \frac{p(h_T|\theta_0,\pi_\phi)}{\frac{1}{L+1}\sum_{\ell=0}^{L} p(h_T|\theta_\ell,\pi_\phi)}\right] =: \mathcal{L}_T^{sPCE}(\pi_\phi, L)$$

$$\partial_\phi \mathcal{I} = \mathbb{E}_{p(\theta_0)p(h_T|\theta_0,\pi_\phi)p(\theta_{1:L})}\left[\log \frac{p(h_T|\theta_0,\pi_\phi)}{\frac{1}{L+1}\sum_0^L p(h_T|\theta_\ell,\pi_\phi)} \cdot \nabla_\phi \log p(h_T|\theta_0,\pi_\phi) + \nabla_\phi\left(\log \frac{\cancel{p(h_T|\theta_0,\pi_\phi)}}{\frac{1}{L+1}\sum_0^L p(h_T|\theta_\ell,\pi_\phi)}\right)\right]$$

$$\mathbb{E}[\nabla_\phi \log p(h_T|\theta_0,\pi_\phi)] = 0$$