Rényi Divergence Variational Inference

Kseniia Petrushina

MIPT, 2023

October 17, 2023

Motivation

Main idea

Approximate inference is a crucial component of modern probabilistic machine learning, as it involves approximating posterior distributions and likelihood functions.
The authors suggest combining existing methods into a unified framework.

Background

Variational inference

Posterior distribution

$$p(\boldsymbol{\theta}|\mathcal{D}, \varphi) = \frac{p(\boldsymbol{\theta}, \mathcal{D}|\varphi)}{p(\mathcal{D}|\varphi)} \tag{1}$$

Variational lower-bound of $p(\mathcal{D}|\varphi)$

$$\mathcal{L}_{VI}(q; \mathcal{D}, \varphi) = \log p(\mathcal{D}|\varphi) - KL[q||p] = \mathbb{E}_q \left[\log \frac{p(\boldsymbol{\theta}, \mathcal{D}|\varphi)}{q(\boldsymbol{\theta})} \right] \tag{2}$$

Rényi's $\alpha - divergence$

$$D_\alpha[p||q] = \frac{1}{\alpha - 1} \log \int p(\boldsymbol{\theta})^\alpha q(\boldsymbol{\theta})^{1-\alpha} d\boldsymbol{\theta} \tag{3}$$

1. continuous and non-decreasing on $\alpha \in \{\alpha : |D_\alpha| < +\infty\}$
2. for $\alpha \notin \{0, 1\}, D_\alpha[p||q] = \frac{\alpha}{1-\alpha} D_{1-\alpha}[p||q] \to D_\alpha[p||q] \leq 0, \alpha < 0$

Variational Rényi bound

Definition

$$\mathcal{L}_\alpha(q; \mathcal{D}, \varphi) = \log p(\mathcal{D}|\varphi) - D_\alpha[q||p] = \frac{1}{1-\alpha}\mathbb{E}_q\left[\log\left(\frac{p(\boldsymbol{\theta}, \mathcal{D}|\varphi)}{q(\boldsymbol{\theta})}\right)^{1-\alpha}\right] \tag{4}$$

Theorem 1

For all $\alpha_+ \in (0, 1)$ and $\alpha_- < 0$

$$\mathcal{L}_\alpha(q; \mathcal{D}) = \lim_{\alpha \to 1} \mathcal{L}_\alpha(q; \mathcal{D}) \leq \mathcal{L}_{\alpha_+}(q; \mathcal{D}) \leq \mathcal{L}_0(q; \mathcal{D}) \leq \mathcal{L}_{\alpha_-}(q; \mathcal{D}) \tag{5}$$

$$\mathcal{L}_0(q; \mathcal{D}) = \log p(\mathcal{D}) \Leftrightarrow \text{supp}(p(\boldsymbol{\theta}|\mathcal{D})) \subseteq \text{supp}(q(\boldsymbol{\theta})) \tag{6}$$

Variational Rényi bound

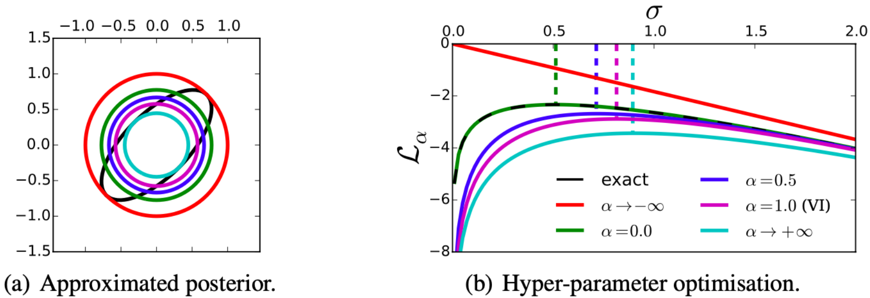
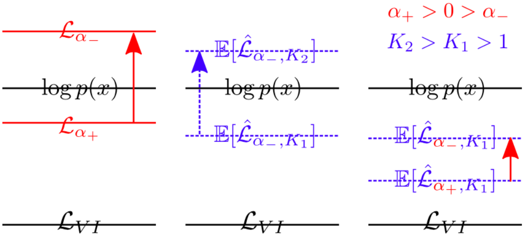


Figure: Mean-Field approximation for Bayesian linear regression.

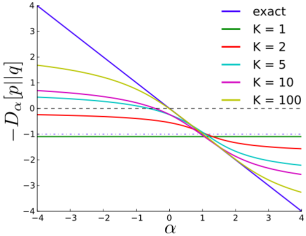
Monte Carlo approximation

Approximation formula

$$\hat{\mathcal{L}}_{\alpha,K}(q;\mathbf{x}) = \frac{1}{1-\alpha}\log\frac{1}{K}\sum_{k=1}^{K}\left[\left(\frac{p(\boldsymbol{\theta}_k,\mathbf{x})}{q(\boldsymbol{\theta}_k|\mathbf{x})}\right)^{1-\alpha}\right], \boldsymbol{\theta}_k \sim q(\boldsymbol{\theta}|\mathbf{x}) \qquad (7)$$



(a) MC approximated VR bounds.



(b) Simulated MC approximations.

Figure: Monte Carlo approximation to the VR bound.

Bayesian neural network

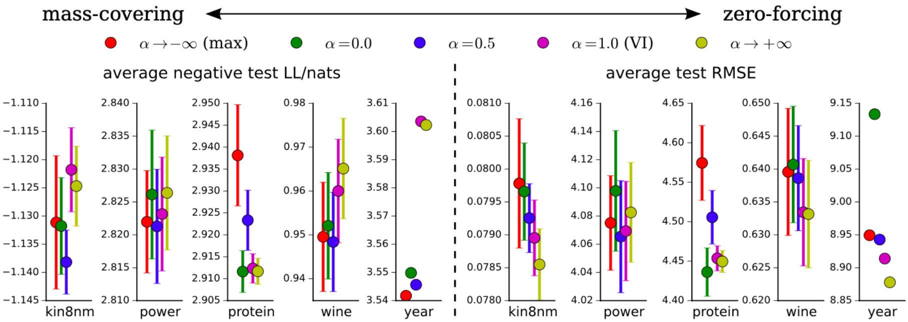


Figure: Test LL and RMSE results for Bayesian neural network regression.

Variational auto-encoder

Dataset	K	**VAE**	**IWAE**	**VR-max**
Frey Face	5	1322.96	**1380.30**	1377.40
(± std. err.)		±10.03	**±4.60**	±4.59
Caltech 101	5	-119.69	**-117.89**	-118.01
Silhouettes	50	-119.61	-117.21	**-117.10**
MNIST	5	-86.47	**-85.41**	-118.01
	50	-86.35	**-84.80**	-117.10
OMNIGLOT	5	-107.62	**-106.30**	-106.33
	50	-107.80	**-104.68**	-105.05

Table: Average Test log-likelihood.

Literature

1. **Main article** Rényi Divergence Variational Inference.