

Stochastic Expectation Propagation

Tue36

Yingzhen Li¹, José Miguel Hernández-Lobato² and Richard E. Turner¹

¹University of Cambridge, ²Harvard University

Goal: approximate the true posterior $q(\theta) \approx p(\theta|\mathcal{D})$

$$p(\theta|\mathcal{D}) \propto p_0(\theta) p(x_1|\theta) p(x_2|\theta) p(x_3|\theta) \approx$$

$$q(\theta) \propto p_0(\theta) f_1(\theta) f_2(\theta) f_3(\theta)$$

idealised

$$p(\theta|\mathcal{D}) \propto p_0(\theta) p(x_1|\theta) p(x_2|\theta) p(x_3|\theta) \approx$$

capture affect of $p(x_3|\theta)$ on posterior

$$p_0(\theta) p(x_1|\theta) p(x_2|\theta) f_3^{new}(\theta)$$

EP

approximate

$$\tilde{p}(\theta) \propto p_0(\theta) f_1(\theta) f_2(\theta) p(x_3|\theta) \approx$$

intractable

$$q_{EP}^{new}(\theta) \propto p_0(\theta) f_1(\theta) f_2(\theta) f_3^{new}(\theta)$$

memory $\mathcal{O}(N)$

Stochastic Expectation Propagation

Tue36

Yingzhen Li¹, José Miguel Hernández-Lobato² and Richard E. Turner¹

¹University of Cambridge, ²Harvard University

Goal: approximate the true posterior $q(\theta) \approx p(\theta|\mathcal{D})$

$$p(\theta|\mathcal{D}) \propto p_0(\theta) p(x_1|\theta) p(x_2|\theta) p(x_3|\theta)$$

$$q(\theta) \propto p_0(\theta) f(\theta)^N$$

capture **average affect** of likelihood terms on posterior

Stochastic EP (N=3)

$$\tilde{p}(\theta) \propto p_0(\theta) f(\theta)^{N-1} p(x_3|\theta)$$

$$q_{SEP}^{new}(\theta) \propto p_0(\theta) f^{new}(\theta)^N$$

EP

$$\tilde{p}(\theta) \propto p_0(\theta) f_1(\theta) f_2(\theta) p(x_3|\theta)$$

$$q_{EP}^{new}(\theta) \propto p_0(\theta) f_1(\theta) f_2(\theta) f_3^{new}(\theta)$$

memory $\mathcal{O}(1)$

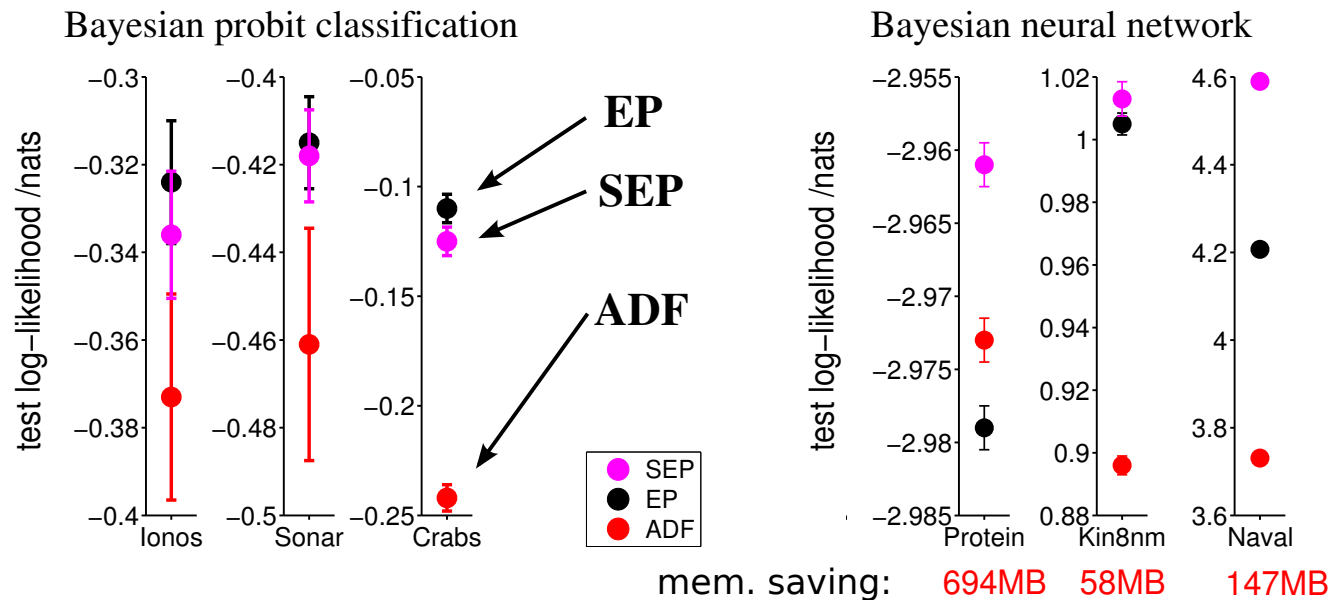
memory $\mathcal{O}(N)$

Stochastic Expectation Propagation

Tue36

Yingzhen Li¹, José Miguel Hernández-Lobato² and Richard E. Turner¹

¹University of Cambridge, ²Harvard University



Performance: **SEP \approx EP, both $>$ ADF**

Memory reduction: **$\mathcal{O}(N) \rightarrow \mathcal{O}(1)$**

(e.g. **~ 65 GB** on YearPredictionMSD dataset)

New results: large-scale GP classification, deep GPs, and interpolating between VB and EP.
(at the Approximate Inference & Black Box Inference workshops)