Kernel-Based Just-In-Time Learning For Passing Expectation Propagation Messages

Wittawat Jitkrittum,1 Arthur Gretton,1 Nicolas Heess, S. M. Ali Eslami, Balaji Lakshminarayanan,1 Dino Sejdinovic2 and Zoltán Szabó1
Gatsby Unit, University College London1 University of Oxford2

Introduction

EP is a widely used message passing based inference algorithm.

Problem: Expensive to compute outgoing from incoming messages.

Goal: Speed up computation by a cheap regression function (message operator):

incoming messages → outgoing message.

Merits:

Efficient online update of the operator during inference.

Uncertainty monitored to invoke new training examples when needed.

Automatic random feature representation of incoming messages.

Expectation Propagation (EP)

Under an approximation that each factor fully factorizes, an outgoing EP message $m_{f\rightarrow i}$ takes the form

$\text{set of } c \text{ variables connected to } f \text{ projected message }$

$\text{projected message}$

$\text{incoming message from } V_i$

Proposed message:

- $q_{f\rightarrow i}(v) = \text{proj} \left[ \int f(y) \prod_{j=1}^{c} m_{f\rightarrow j}(y_j) \, dV_j(y_j) \right] m_{V_i\rightarrow f}(v_i) =: q_{f\rightarrow V_i}(v_i)$

$\text{projection onto exponential family}$

Kernel on Incoming Messages

Propose to incrementally learn during inference a kernel-based EP message operator (distribution-to-distribution regression)

$\text{for any factor } f \text{ that can be sampled.}$

- Product distribution of incoming messages: $r = \times_{i \in I} r_i$, $s = \times_{i \in J} s_i$

- Mean embedding of $r$: $\mu_r = \mathbb{E}_{r \sim F} \left[ r \right]$

- Gaussian kernel on (product) distributions. Two-staged random feature approx.: $

\kappa(r, s) = \exp \left[ -\frac{\|\mu_r - \mu_s\|}{2\gamma^2} \right] \exp \left[ -\frac{\|\phi(r) - \phi(s)\|}{2\gamma^2} \right] \approx \kappa(r) \kappa(s)$

Message Operator: Bayesian Linear Regression

Input: $X = (x_1, \cdots, x_N)$: $N$ training incoming messages represented as random feature vectors.

Output: $Y = (\mathbb{E}_r \mu_{\mathbb{E}_r} u(y)) \cdots (\mathbb{E}_s \mu_{\mathbb{E}_s} u(y)) \in \mathbb{R}^{1 \times N}$: expected sufficient statistics of outgoing messages.

Inexpensive online updates of posterior mean and covariance.

Bayesian regression gives prediction and predictive variance.

If predictive variance > threshold, query the importance sampling oracle.

Two-Staged Random Features

In: $\mathcal{F}(k)$: Fourier transform of $k$, $D_{in}$: #inner features, $D_{out}$: #outer features,

$K_{gau}$: Gaussian kernel on $\mathbb{R}^{D_{out}}$.

Out: Random features $\psi(r) \in \mathbb{R}^{D_{out}}$

1. Sample $\{w_i\}_{i=1}^{D_{in}} \sim i.i.d \mathcal{F}(k), \{b_i\}_{i=1}^{D_{in}} \sim i.i.d \mathcal{U}[0, 2\pi]$.

2. $\phi(r) = \mathbb{E}_r \left[ \mathbb{E}_s \cos \left[ w_i^\top x + b_i \right] \right] \in \mathbb{R}^{D_{in}}$.

3. $\psi(r) = \mathbb{E}_r \left[ \mathbb{E}_s \phi(r) \right] \in \mathbb{R}^{D_{in}}$.

Experiment 1: Uncertainty Estimates

- Approx: $f(p|x) = \delta \left[ p - \frac{1}{\sqrt{\text{uncertainty}}} \right]$

- Training messages collected from 20 EP runs on toy data.

- #Random features: $D_{in} = 300$ and $D_{out} = 500$.

Experiment 2: Real Data

- Binary logistic regression. Sequentially present 4 real datasets to the operator.

- Diverse distributions of incoming messages.

- Sampling + KJIT = proposed KJIT with an importance sampling oracle.

- KJIT operator can adapt to the change of input message distributions.

Experiment 3: Compound Gamma Factor

Infer posterior of the precision $\tau$ of $x \sim N(\mu, \tau^{-1})$ from observations $\{x_i\}_{i=1}^{N}$:

$y_i \sim \text{Gamma}\left(\gamma_i; \tau, \mu \right)$

Prec.: $

\nu \sim \text{Gamma}\left(\tau; \sigma_2, \Sigma_2 \right)$

$\tau \sim \text{Gamma}\left(\gamma_2; \tau, \Sigma_2 \left(\nu^{-1} \right)^{-1} \right)$

$I_i \sim \text{Gamma}\left(\gamma_3; \nu, \Sigma_3 \right)$

Inferred by Infer.NET + KJIT

- Infer.NET + KJIT = proposed KJIT with a hand-crafted factor as oracle.

- Inference quality: as good as hand-crafted factor; much faster.

Contact: wittawat@gatsby.ucl.ac.uk  We thank the Gatsby Charitable Foundation for the financial support.