Variational Gaussian Process Models without Matrix Inverses

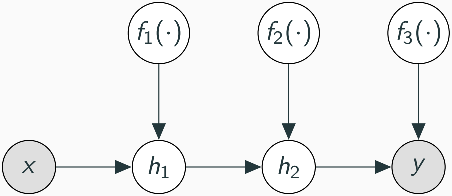
Mark van der Wilk, ST John, Artem Artemev, James Hensman

December 8, 2019

PROWLER.io

Gaussian processes as building blocks

Q: Can Gaussian processes be building blocks
for Bayesian deep learning?



Variational inference is **general** and **actually works** in GP models:

- the ELBO is tight enough for **hyperparameter selection**.
- the variational posterior predicts with **non-parametric error bars**.

Scalability

Cost of minibatch iteration: $\mathcal{O}(BM^2 + M^3)$
for batch size B and model capacity M.

$$\mathcal{L} = \sum_{n=1}^{N} \mathbb{E}_{q(f(\mathbf{x}_n))}[\log p(y_n \mid f(\mathbf{x}_n)] - \mathrm{KL}[q(f(\cdot))||p(f(\cdot))]$$

$$q(f(\mathbf{x}_n)) = \mathcal{N}\left(\mathbf{k}_{\mathbf{u}f_n}^{\top}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{m}, k_{f_n f_n} - \underbrace{\mathbf{k}_{\mathbf{u}f_n}^{\top}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}f_n}}_{\text{difficult}} + \mathbf{k}_{\mathbf{u}f_n}^{\top}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{S}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}f_n}\right)$$

Scalability

Cost of minibatch iteration: $\mathcal{O}(BM^2 + M^3)$
for batch size B and model capacity M.

$$\mathcal{L} = \sum_{n=1}^{N} \mathbb{E}_{q(f(\mathbf{x}_n))}[\log p(y_n \mid f(\mathbf{x}_n)] - \mathrm{KL}[q(f(\cdot))||p(f(\cdot))]$$

$$q(f(\mathbf{x}_n)) = \mathcal{N}\left(\mathbf{k}_{\mathbf{u}f_n}^{\top}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{m}, k_{f_n f_n} - \underbrace{\mathbf{k}_{\mathbf{u}f_n}^{\top}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}f_n}}_{\text{difficult}} + \mathbf{k}_{\mathbf{u}f_n}^{\top}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{S}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}f_n}\right)$$

Q: Can Gaussian processes be building blocks
for Bayesian deep learning?

Scalability

Cost of minibatch iteration: $\mathcal{O}(BM^2 + M^3)$
for batch size B and model capacity M.

$$\mathcal{L} = \sum_{n=1}^{N} \mathbb{E}_{q(f(\mathbf{x}_n))}[\log p(y_n \mid f(\mathbf{x}_n)] - \mathrm{KL}[q(f(\cdot))||p(f(\cdot))]$$

$$q(f(\mathbf{x}_n)) = \mathcal{N}\left(\mathbf{k}_{\mathbf{u}f_n}^\top \mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{m}, k_{f_n f_n} - \underbrace{\mathbf{k}_{\mathbf{u}f_n}^\top \mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}f_n}}_{\text{difficult}} + \mathbf{k}_{\mathbf{u}f_n}^\top \mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{S}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}f_n}\right)$$

Q: Can Gaussian processes be building blocks
for Bayesian deep learning?

A: As long as they need a matrix inverse...
probably not.

Inverse-free Gaussian Processes

1. New variational posterior:

   Hensman et al [2013] posterior:

   $$\mathcal{N}\Big(\mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{m},\ k_{f_nf_n} - \mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}f_n} + \qquad \mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{S}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}f_n}\Big)$$

   Our posterior:

   $$\mathcal{N}\Big(\mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{m},\ k_{f_nf_n} + \mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{T}\mathbf{K}_{\mathbf{uu}}\mathbf{T}\mathbf{k}_{\mathbf{u}f_n} - 2\mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{T}\mathbf{k}_{\mathbf{u}f_n} + \qquad \mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{S}'\mathbf{k}_{\mathbf{u}f_n}\Big)$$

Inverse-free Gaussian Processes

1. New variational posterior:

   Hensman et al [2013] posterior:

   $$\mathcal{N}\big(\mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{m},\ k_{f_nf_n} - \mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}f_n} + \qquad \mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{S}\mathbf{K}_{\mathbf{uu}}^{-1}\mathbf{k}_{\mathbf{u}f_n}\big)$$

   Our posterior:

   $$\mathcal{N}\big(\mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{m},\ k_{f_nf_n} + \mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{T}\mathbf{K}_{\mathbf{uu}}\mathbf{T}\mathbf{k}_{\mathbf{u}f_n} - 2\mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{T}\mathbf{k}_{\mathbf{u}f_n} + \qquad \mathbf{k}_{\mathbf{u}f_n}^{\intercal}\mathbf{S}'\mathbf{k}_{\mathbf{u}f_n}\big)$$

2. New variational bound also a function of $\mathbf{T}$, with property:

   $$\mathbf{T}^* = \underset{\mathbf{T}}{\operatorname{argmax}}\,\mathcal{L}(\ldots, \mathbf{T}) = \mathbf{K}_{\mathbf{uu}}^{-1}$$

Inverse-free Gaussian Processes

1. New variational posterior:

   Hensman et al [2013] posterior:

   $$\mathcal{N}\big(\mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{m},\ k_{f_n f_n} - \mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{K}_{\mathbf{uu}}^{-1} \mathbf{k}_{\mathbf{u}f_n} + \mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{K}_{\mathbf{uu}}^{-1} \mathbf{S} \mathbf{K}_{\mathbf{uu}}^{-1} \mathbf{k}_{\mathbf{u}f_n}\big)$$

   Our posterior:

   $$\mathcal{N}\big(\mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{m},\ k_{f_n f_n} + \mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{T} \mathbf{K}_{\mathbf{uu}} \mathbf{T} \mathbf{k}_{\mathbf{u}f_n} - 2\mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{T} \mathbf{k}_{\mathbf{u}f_n} + \mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{S}' \mathbf{k}_{\mathbf{u}f_n}\big)$$

2. New variational bound also a function of $\mathbf{T}$, with property:

   $$\mathbf{T}^* = \underset{\mathbf{T}}{\operatorname{argmax}}\, \mathcal{L}(\ldots, \mathbf{T}) = \mathbf{K}_{\mathbf{uu}}^{-1}$$

3. Stochastic estimate of bound costs $\mathcal{O}(BM^2)$

Inverse-free Gaussian Processes

1. New variational posterior:

   Hensman et al [2013] posterior:

   $$\mathcal{N}\Big(\mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{m},\ k_{f_n f_n} - \mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{K}_{\mathbf{uu}}^{-1} \mathbf{k}_{\mathbf{u}f_n} + \qquad \mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{K}_{\mathbf{uu}}^{-1} \mathbf{S} \mathbf{K}_{\mathbf{uu}}^{-1} \mathbf{k}_{\mathbf{u}f_n}\Big)$$

   Our posterior:

   $$\mathcal{N}\Big(\mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{m},\ k_{f_n f_n} + \mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{T} \mathbf{K}_{\mathbf{uu}} \mathbf{T} \mathbf{k}_{\mathbf{u}f_n} - 2\mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{T} \mathbf{k}_{\mathbf{u}f_n} + \qquad \mathbf{k}_{\mathbf{u}f_n}^\intercal \mathbf{S}' \mathbf{k}_{\mathbf{u}f_n}\Big)$$

2. New variational bound also a function of $\mathbf{T}$, with property:

   $$\mathbf{T}^* = \underset{\mathbf{T}}{\operatorname{argmax}}\, \mathcal{L}(..., \mathbf{T}) = \mathbf{K}_{\mathbf{uu}}^{-1}$$

3. Stochastic estimate of bound costs $\mathcal{O}(BM^2)$

See more at our poster #49!