Sequential Learning in GPs with Memory and Bayesian Leverage Score




Prakhar Verma
Aalto University

Paul E. Chang
Aalto University

Arno Solin
Aalto University

Mohammad Emtiyaz Khan
RIKEN Center for AI Project

RIKEN Aalto University

Continual Lifelong Learning Workshop – ACML 2022

Sequential Learning in GPs

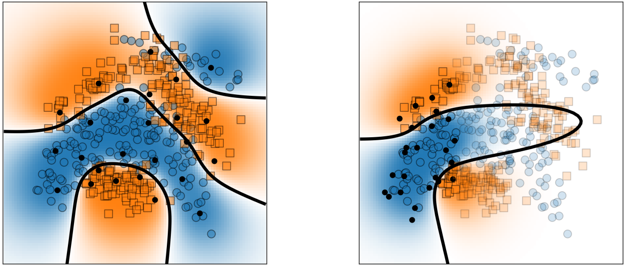
- Sequential decision-making problems like
  - Bayesian Optimization
  - Reinforcement learning
- In offline setting, SVGP is a popular model but fails in sequential setting with limited past data.

Sequential Learning in GPs

- Sequential decision-making problems like
  - Bayesian Optimization
  - Reinforcement learning
- In offline setting, SVGP is a popular model but fails in sequential setting with limited past data.

Sequential Learning in GPs

- Sequential decision-making problems like
  - Bayesian Optimization
  - Reinforcement learning
- In offline setting, SVGP is a popular model but fails in sequential setting with limited past data.



VI in GPs

- SVGP ELBO:

$$\sum_{i\in(\mathcal{D}_{\text{old}}\cup\mathcal{D}_{\text{new}})} \mathbb{E}_{q_{\mathbf{u}}(f_i)}[\log p(y_i \mid f_i)] - \mathbb{D}_{KL}[q(\mathbf{u}) \,\|\, p_\theta(\mathbf{u})]$$

- Approximated ELBO with memory:

$$\sum_{i\in\mathcal{D}_{\text{new}}} \mathbb{E}_{q_{\mathbf{u}}(f_i)}[\log p(y_i \mid f_i)] + S\sum_{i\in\mathcal{M}} \mathbb{E}_{q_{\mathbf{u}}(f_i)}[\log p(y_i \mid f_i)] + \underbrace{\log \mathcal{Z} - \mathbb{E}_{q(\mathbf{u})}[\log \mathrm{N}(\tilde{\mathbf{y}} \mid \mathbf{u}, \tilde{\Sigma})]}_{\mathbb{D}_{KL}[q(\mathbf{u}) \,\|\, p_\theta(\mathbf{u})]}$$

- Memory selection technique?

VI in GPs

- SVGP ELBO:

$$\sum_{i \in (\mathcal{D}_{\text{old}} \cup \mathcal{D}_{\text{new}})} \mathbb{E}_{q_{\mathbf{u}}(f_i)}[\log p(y_i \mid f_i)] - \mathbb{D}_{KL}[q(\mathbf{u}) \,\|\, p_\theta(\mathbf{u})]$$

- Approximated ELBO with memory:

$$\sum_{i \in \mathcal{D}_{\text{new}}} \mathbb{E}_{q_{\mathbf{u}}(f_i)}[\log p(y_i \mid f_i)] + S \sum_{i \in \mathcal{M}} \mathbb{E}_{q_{\mathbf{u}}(f_i)}[\log p(y_i \mid f_i)] + \underbrace{\log \mathcal{Z} - \mathbb{E}_{q(\mathbf{u})}[\log \mathrm{N}(\tilde{\mathbf{y}} \mid \mathbf{u}, \tilde{\boldsymbol{\Sigma}})]}_{\mathbb{D}_{KL}[q(\mathbf{u}) \,\|\, p_\theta(\mathbf{u})]}$$

- Memory selection technique?

VI in GPs

- SVGP ELBO:

$$\sum_{i \in (\mathcal{D}_{\text{old}} \cup \mathcal{D}_{\text{new}})} \mathbb{E}_{q_{\mathbf{u}}(f_i)}[\log p(y_i \mid f_i)] - \mathbb{D}_{KL}[q(\mathbf{u}) \parallel p_\theta(\mathbf{u})]$$

- Approximated ELBO with memory:

$$\sum_{i \in \mathcal{D}_{\text{new}}} \mathbb{E}_{q_{\mathbf{u}}(f_i)}[\log p(y_i \mid f_i)] + S \sum_{i \in \mathcal{M}} \mathbb{E}_{q_{\mathbf{u}}(f_i)}[\log p(y_i \mid f_i)] + \underbrace{\log \mathcal{Z} - \mathbb{E}_{q(\mathbf{u})}[\log \mathrm{N}(\tilde{\mathbf{y}} \mid \mathbf{u}, \tilde{\mathbf{\Sigma}})]}_{\mathbb{D}_{KL}[q(\mathbf{u}) \parallel p_\theta(\mathbf{u})]}$$

- Memory selection technique?

Bayesian Leverage Score

- Ridge leverage score, $\mathrm{diag}(\mathbf{K_{xx}}(\mathbf{K_{xx}} + \lambda \mathbf{I})^{-1})$.
- Bayesian leverage score (BLS),

$$h_i^{\text{bls}} := \left[\mathbf{K_{xx}}(\mathbf{K_{xx}} + \mathrm{diag}(\beta_*^{-1}))^{-1}\right]_{ii} = \beta_i^* v_{f,i,i}^*.$$

- BLS score indicates how difficult the example was for the model.

Bayesian Leverage Score

- Ridge leverage score, $\mathrm{diag}(\mathbf{K_{xx}}(\mathbf{K_{xx}} + \lambda\,\mathbf{I})^{-1})$.
- Bayesian leverage score (BLS),

$$h_i^{\text{bls}} := \left[\mathbf{K_{xx}}(\mathbf{K_{xx}} + \mathrm{diag}(\beta_*^{-1}))^{-1}\right]_{ii} = \beta_i^* v_{f,i,i}^*.$$

- BLS score indicates how difficult the example was for the model.

Bayesian Leverage Score

- Ridge leverage score, $\mathrm{diag}(\mathbf{K_{xx}}(\mathbf{K_{xx}} + \lambda\,\mathbf{I})^{-1})$.
- Bayesian leverage score (BLS),

$$h_i^{\text{bls}} := \left[\mathbf{K_{xx}}(\mathbf{K_{xx}} + \mathrm{diag}(\boldsymbol{\beta}_*^{-1}))^{-1}\right]_{ii} = \beta_i^* v_{f,i,i}^*.$$

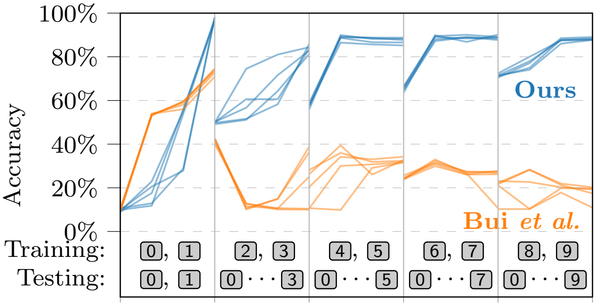
- BLS score indicates how difficult the example was for the model.

BLS: Split MNIST

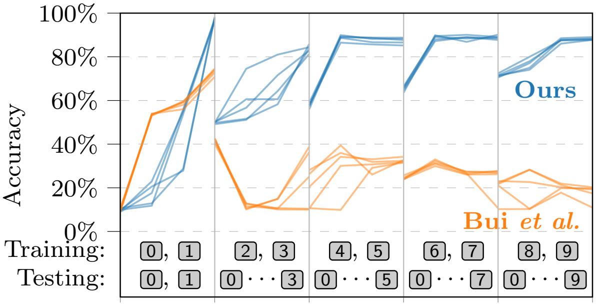
High BLS

Low BLS

Split-MNIST


100%
80%
60%
40%
20%
0%
Accuracy
Ours
Bui et al.
Training:
Testing:
0, 1
2, 3
4, 5
6, 7
8, 9
0, 1
0 · · · 3
0 · · · 5
0 · · · 7
0 · · · 9

Split-MNIST



10% memory size

Thanks!