Article presentation

A. Ndiaye E. Songo

Bayesian Learning Rule

Motivation Principle

Applications o the Bayesian Learning Rule

The special case of ridge regression

Conclusion an Discussion

# The Bayesian Learning Rule (2022)

M. E. Khan & H. Rue

Aminata Ndiaye & Edimah Songo M2 MASH

#### Outline

Article presentation

**Dauphine** | PSL₩

A. Ndiaye E. Songo

Bayesian Learning Rul

Motivation Principle

Applications the Bayesian Learning Rule

The special case o

Conclusion and Discussion

- Bayesian Learning Rule
  - Motivation
  - Principle
- Applications of the Bayesian Learning Rule
  - The special case of ridge regression
- 3 Conclusion and Discussion

# Bayesian Learning Rule

Article presentation

**Dauphine** | PSL₩

A. Ndiaye E. Songo

Bayesian Learning Ru

Motivation Principle

Applications of the Bayesian Learning Rule

The special case o ridge regression

Conclusion and Discussion

#### **Motivation**

- Create a unified framework to derive established and new algorithms
- Improve existent algorithms (convergence speed)
- Find new algorithms

#### How we proceed?

- Optimize a Bayesian objective function
- Finding the posterior distribution of the parameters of interest
- Taking more information into account by using the natural gradient

Bayesian Learning Rule

Principle

Applications the Bayesian Learning Rule

The special case o ridge regression

Conclusion and Discussion

### **Bayesian Objective**

$$q_*( heta) = rg \min_{oldsymbol{q}( heta)} \mathbb{E}_{oldsymbol{q}} \left[ \sum_{i=1}^N \ell(y_i, f_{ heta}(x_i)) 
ight] + \mathbb{D}_{oldsymbol{\mathcal{KL}}}[oldsymbol{q}( heta) \parallel oldsymbol{p}( heta)]$$

### **Setting**

- ullet  $q\in\mathcal{Q}$  a set of regular and minimal exponential family
- $\lambda_{t+1} \leftarrow \lambda_t \rho_t \tilde{\nabla}_{\lambda} \left[ \mathbb{E}_{q_t}[\bar{\ell}(\theta)] \mathcal{H}(q_t) \right]$

### **Steps**

- Choice of posterior approximation (here in the exponential family)
- Choice of approximation method for natural gradient (ex: Delta Method, etc.)

Quadratic loss for some penalty term  $\delta > 0$ 

$$\bar{\ell}(\theta) = \frac{1}{2} (y - X\theta)^T (y - X\theta) + \frac{1}{2} \delta \theta^T \theta$$
$$\theta^* = (X^T X + \delta I)^{-1} X^T y$$

## Natural gradients in Ridge Regression

- Candidate posterior:  $\mathcal{N}(m, S)$
- $\mu = E[T(\theta)]$
- Natural gradients:

$$\tilde{\nabla}_{\mu^{(1)}} E_q[\bar{\ell}(\theta)] = -X^T y$$
 and  $\tilde{\nabla}_{\mu^{(2)}} E_q[\bar{\ell}(\theta)] = \frac{1}{2} (X^T X + \delta I) \mu^{(2)}$ 

• Solution:  $\theta^* = m^* = (S^*)^{-1}X^T y = (X^TX + \delta I)^{-1}X^T y$ 

#### Conclusion

Article presentation

Dauphine | PSL®

A. Ndiaye E. Songo

Bayesian Learning Ru Motivation

Applications the Bayesian Learning Rule

The special case o ridge regression

Conclusion and Discussion

#### **Article's takeaways**

- Research possibilities new for state-of-the-art algorithms
- Only two choices are required (posterior and natural gradient approximations)

#### **Drawbacks**

- The article was at times difficult to understand
- Algorithms not necessarily optimal
  - Restriction to exponential families
  - Difficult to compute gradients
  - Topological instability

#### Potential solution to some issues

New research: Lie-Group Bayesian Learning Rule (2023)