Bayesian C-DF

13.12.2018

Academic Seminar

Wonjai Lee

Index

About B C-DF

Online Learning ⟷ Batch Learning

- Sequential
- Simultaneously

- Subsequently
- Calculate at Once

About B C-DF

Key Point of online learning

- By replacing the true Posterior distribution

1. Update approximate posterior
2. Optimal projection into parametric family (Choosing it to be Gaussian)

- Simultaneously

About B C-DF

In terms of Epistemic Uncertainty

$$p(\theta|D_t) = \frac{p(\theta)P(D_t|\theta)}{\int d\theta' p(\theta')P(D_t|\theta')}.$$

About B C-DF

MCMC(Markov Chain Monte Carlo)

$$p(\theta|D_t) = \frac{p(\theta)P(D_t|\theta)}{\boxed{\int d\theta' p(\theta')P(D_t|\theta')}} \Rightarrow \text{Hard to Calculate}$$

About B C-DF

Advantages

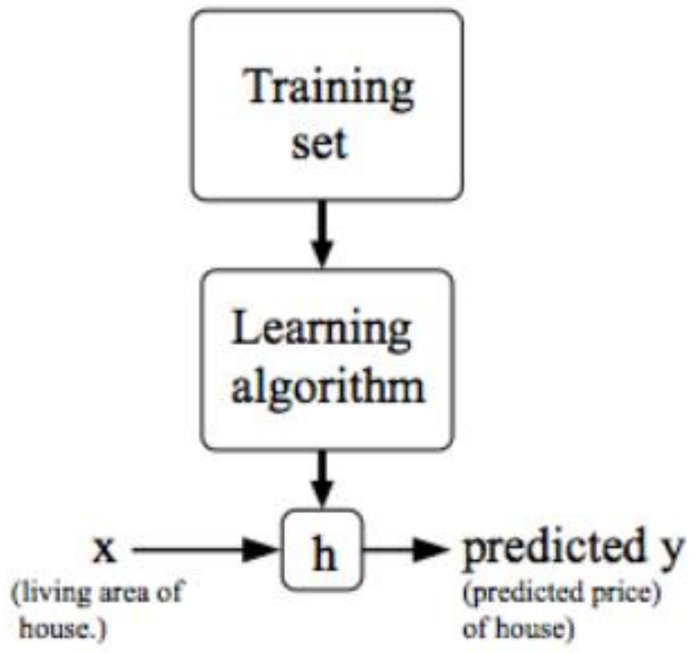
1. Simultaneously
2. Serves numerous desirable feature
3. High Dimensional compressed regression

About B C-DF

Existed Methods

- ADF(Assumed Density Filtering)
- EP(Expectation Propagation)
- PL(Particle Learning)
- SMC(Sequential Monte Carlo)

Linear Regression



Linear Regression

Cost function

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2 = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$$

Minimize the cost function

Linear Regression

Gradient Descent

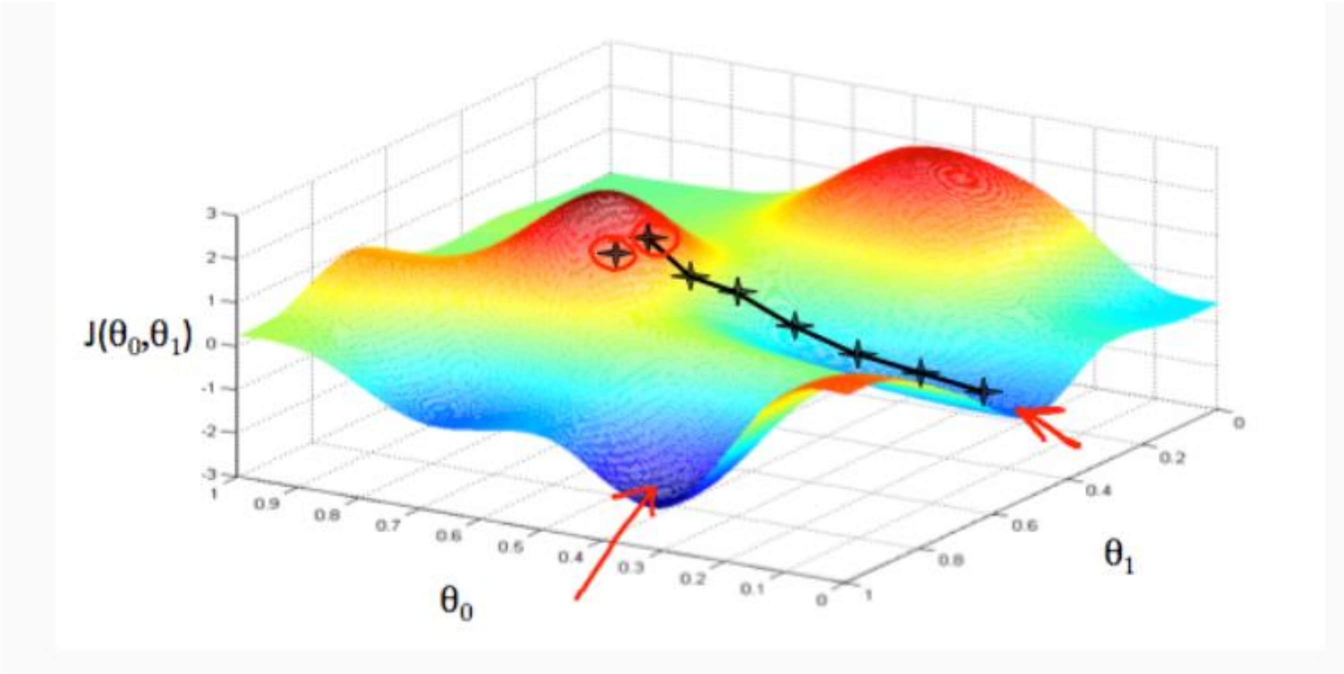
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

where

j=0,1 represents the feature index number.

Trace the minimum value of J

Linear Regression



Example of B C-DF in Linear Regression

For Gaussian Error Model

$y \in \Re$ given an associated *p*-dimensional predictor

$\boldsymbol{x} \in \Re^p$ modeled in the linear regression setting as
y ~ N*(**xβ**, σ2).*

$$\boldsymbol{y}^t = \boldsymbol{X}^t\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n).$$

Example of B C-DF in Linear Regression

A standard Bayesian analysis proceeds by assigning conjugate prior *(β, σ2)* ~ N*(*0*, Ip)* × IG*(a, b)* with

$$S_t^{XX} = S_{t-1}^{XX} + X^{t'} X^t$$

$$S_t^{XY} = S_{t-1}^{XY} + X^{t'} y^t$$

$$S_t^{YY} = S_{t-1}^{YY} + y^{t'} y^t$$

Parameters of distribution that enable online inference using MCMC

Example of B C-DF in Linear Regression

Namely with Standard Conjugation Priors

$$\sigma^2 | \boldsymbol{\beta}, \boldsymbol{D}^{(t)} \sim \mathrm{IG}(a_t, b_t), \quad a_t = a + nt/2,$$

$$b_t = b + \frac{1}{2}\left(S_t^{YY} - 2\boldsymbol{\beta}' \boldsymbol{S}_t^{XY} + \boldsymbol{\beta}' \boldsymbol{S}_t^{XX} \boldsymbol{\beta}\right)$$

$$\boldsymbol{\beta} | \sigma^2, \boldsymbol{D}^{(t)} \sim N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), \quad \boldsymbol{\Sigma}_t = \left(\boldsymbol{S}_t^{XX}/\sigma^2 + \boldsymbol{I}_p\right)^{-1}$$

$$\boldsymbol{\mu}_t = \boldsymbol{\Sigma}_t \boldsymbol{S}_t^{XY}/\sigma^2.$$

Example of B C-DF in Linear Regression

Procedure

0. C-DF begins by defining a partition of modeled parameters

$$\Theta_{\mathcal{G}_1} = \{\beta\}$$
$$\Theta_{\mathcal{G}_2} = \{\sigma^2\}.$$

Example of B C-DF in Linear Regression

Procedure

1. Observe data *Dt* at time *t*. At *t* = 1 initialize all parameters at some default values.

(e.g., $\boldsymbol{\beta} = \mathbf{0}, \sigma^2 = 1$)

Example of B C-DF in Linear Regression

Procedure

2. Define first parameter and update

$$C_1^{(t)} = \{C_{1,1}^{(t)}, C_{1,2}^{(t)}\}$$

$$C_{1,1}^{(t)} = C_{1,1}^{(t-1)} + X^{t'} X^t / \hat{\sigma}_{t-1}^2$$

$$C_{1,2}^{(t)} = C_{1,2}^{(t-1)} + X^{t'} y^t / \hat{\sigma}_{t-1}^2$$

Example of B C-DF in Linear Regression

Procedure

3. Draw *S* samples from the approximate Gibbs full conditional

$$\boldsymbol{\beta} | \hat{\sigma}^2_{t-1}, \boldsymbol{C}_1^{(t)} \sim \mathrm{N}(\hat{\boldsymbol{\mu}}_t, \boldsymbol{\Sigma}_t)$$

$$(\hat{\boldsymbol{\Sigma}}_t = (\boldsymbol{C}_{1,1}^{(t)} + \boldsymbol{I}_p)^{-1}, \hat{\boldsymbol{\mu}}_t = \hat{\boldsymbol{\Sigma}}_t \boldsymbol{C}_{1,2}^{(t)})$$

Example of B C-DF in Linear Regression

Procedure

4. Define second parameter and update

$$\mathbf{C}_2^{(t)} = \{\mathbf{C}_{2,1}^{(t)}, \mathbf{C}_{2,2}^{(t)}, C_{2,3}^{(t)}\}$$

$$\mathbf{C}_{2.1}^{(t)} = \mathbf{C}_{2,1}^{(t-1)} + \hat{\boldsymbol{\beta}}'_{t-1}\boldsymbol{X}^{t'}\boldsymbol{X}^{t}\hat{\boldsymbol{\beta}}_{t-1},$$

$$\mathbf{C}_{2,2}^{(t)} = \mathbf{C}_{2,2}^{(t-1)} + \boldsymbol{\beta}_{t-1}\boldsymbol{X}^{t'}\boldsymbol{y}^{t}$$

$$\mathbf{C}_{2,3}^{(t)} = \mathbf{C}_{2,3}^{(t-1)} + \boldsymbol{y}^{t'}\boldsymbol{y}^{t}.$$

Example of B C-DF in Linear Regression

Procedure

5. Draw S samples from the approximate Gibbs full conditional

$$\sigma^2 | \hat{\boldsymbol{\beta}}_{t-1}, \boldsymbol{C}_2^{(t)} \sim \mathrm{IG}(a_t', b + (C_{2,3}^{(t)} - 2\boldsymbol{C}_{2,2}^{(t)} + \boldsymbol{C}_{2,1}^{(t)})/2);$$

Example of B C-DF in Linear Regression

Results

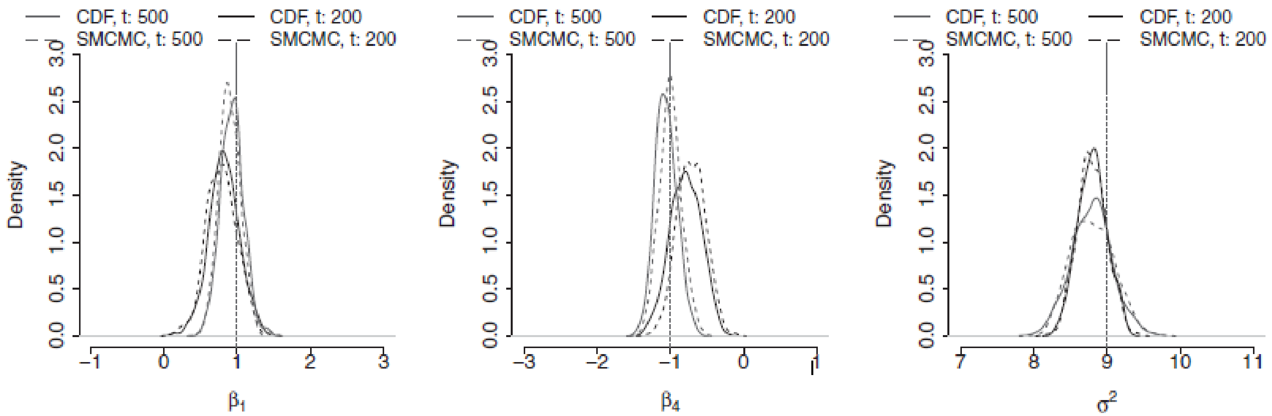


Figure 1. Kernel density estimates for posterior draws using the C-DF algorithm and S-MCMC at $t = 200, 500$.

Example of B C-DF in Linear Regression

Results

Table 1. Inferential performance for C-DF and S-MCMC for parameters of interest. Coverage and length are based on 95% credible intervals averaged over all β_j ($j = 1, \ldots, 5$) and over 10 replications. We report the time taken to draw $M = 500$ MCMC samples with the arrival of each data shard. Complexity of sampling M MCMC samples are also reported under flops. MSE and their standard error are reported at different time points.

					$\text{MSE} = \sum_{j=1}^{p}(\hat{\beta}_t - \beta_0)^2/p$		
	Avg. Cov. β	Length	Time (sec)	flops	$t = 200$	$t = 500$	$t = 1000$
C-DF	0.96	$0.49_{0.01}$	$95_{4.12}$	M	$0.10_{0.001}$	$0.014_{0.001}$	$0.003_{0.001}$
S-MCMC	1.0	$0.49_{0.01}$	$119.4_{4.64}$	Mp^3	$0.06_{0.001}$	$0.005_{0.001}$	$0.003_{0.001}$

Result

- Facilitating efficient online Bayesian inference by adapting MCMC obtained by propagating surrogate statistics as new data arrive.

- Eliminates the need to store or process the entire data at once which often results in large computational savings.

- Being accompanied with good runtime, memory and sampling efficiency improvements over various state-of-the-art competitors.

Thank you!