Meeting

2019/07/01

1. Summer research credit program - Planned first week schedule

1) Literature Search
   1. Bayesian Neural Network + Video Interpolation
      - Not practical generally
      - No gain in generalization performance but additional computational costs.
        $\Rightarrow$There are some reasons if nobody does it.
      - However, we can diagnose model behavior => better model interpretability
        $\Rightarrow$Useful?
2) Research environment setup
   1. Getting friendly with Lab server + Docker
      - Created own docker environment with docker build
        $\Rightarrow$ Better productivity

1. Literature Search

1) Variational Inference: A review for statisticians
2) Uncertainty estimations by softplus normalization in Bayesian Convolutional Neural Networks with Variational Inference

Type of Estimators(Mathematical Stat)

- **Method of Moments(MOM)**: need complex calculations often, hard integration => not practical, don't use in practice.
- **Maximum Likelihood Estimation(MLE)**: Basically used in ML. Efficient, nice theoretical guarantees.
- **Maximum a Posteriori(MAP)**: Only calculates point estimate in full Bayesian. Similar computation with MLE.
- **Full Bayesian**: Need to calculate integration or approximation. Can model posterior distribution $p(\theta|D)$ (Uncertainty estimation). But, typically lager amount of computation.

Bayesian Methods

$$Posterior = \frac{Prior \times Likelihood}{Evidence}$$

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{\int_{\theta'} p(\theta')p(D|\theta')d\theta'} = \frac{p(\theta)\prod_{i=1}^{n} p(y_i|x_i,\theta)}{\int_{\theta'} p(\theta')\prod_{i=1}^{n} p(y_i|x_i,\theta')d\theta'}$$

*Substitute Summation for integration for discrete RV

$\Rightarrow$We can calculate Posterior distribution

Intractable Posterior distribution

- Current integration algorithm has time complexity of $O(n^p)$ for number of variables P(Monte carlo, etc)
- Calculating Evidence term $\int_{\theta'} p(\theta') \prod_{i=1}^{n} p(y_i|x_i, \theta') d\theta'$ is generally intractable.
- But nowadays, efficient evidence approximation framework called "Variational Inference" is introduced and widely used.

Variational Inference

$$\begin{aligned}\theta^{opt} &= \arg\min_{\theta} \mathrm{KL}\left[q_\theta(w|\mathcal{D}) \| p(w|\mathcal{D})\right] \\ &= \arg\min_{\theta} \mathrm{KL}\left[q_\theta(w|\mathcal{D}) \| p(w)\right] - \mathbb{E}_{q(w|\theta)}\left[\log p(\mathcal{D}|w)\right] + \log p(\mathcal{D})\end{aligned} \tag{1}$$

where

$$\mathrm{KL}\left[q_\theta(w|\mathcal{D}) \| p(w)\right] = \int q_\theta(w|\mathcal{D}) \log \frac{q_\theta(w|\mathcal{D})}{p(w)} dw. \tag{2}$$

- Transform approximation problem to optimization problem
- However, another approximation problem arises.
- But, easier and accurate approximation with standard Monte Carlo

1. Literature Search

1) Variational Inference: A review for statisticians
2) Uncertainty estimations by softplus normalization in Bayesian Convolutional Neural Networks with Variational Inference

Bayesian Neural Networks?

Bayes by Backprop

- Standard Backprop and update for NN:
  1. Use chain rule to calculate derivative of each parameters
  2. Apply gradient-based optimization algorithm.

Bayes by Backprop

- Bayes by backprop(probabilistic backprop):
  - Also called local reparameterization trick
  - Algorithm for gaussian distribution modeling
    1. Sample from gaussian ϵ ~ N(0, 1)
    2. Multiply s.d parameter σ and add mean parameter μ
    3. Use chain rule to calculate gradient for each parameter.
    4. Apply gradient-based optimization algorithm

    => Modeling for arbitrary Gaussian distribution is possible!

Other personal improvements

1. PRML ch2 exercises(half)
2. PRML ch3 offline meeting with SKKU undergrads & grads

- Participated SKKU statistic academy "P-SAT" home coming day(7/29): Networking with stat/CS graduate students, newly employed data analyst

Other discussions

- ICCV 2019 in seoul
- In-lab Study