[211112] SPARK LAB MEETING

Variational Inference

김정훈

MAP(Maximum A Posteriori estimation)

- Bayes' theorem

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} = \frac{P(X|\theta)P(\theta)}{\int P(X|\theta)\,P(\theta)d\theta} \longleftarrow$$ **marginalize**

- MAP : parameter θ that maximize posterior

$$\hat{\theta}_{MAP} = argmax_{\theta} p(\theta|x_n)$$

- MAP : observations, likelihood probability, prior distribution, random variable θ
- MLE : observations , likelihood probability(θ)
- Estimate deterministic parameter θ and frizzing parameter in **new data**

- Posterior Predictive Distribution (parametric model)
  - Consider all cases

$$p(x_n|X) = \int P(x_n|\theta, X)\,P(\theta|X)d\theta \longleftarrow$$ **marginalize**

Spark-lab youtube, [SCS4049] Inclass 18 | Maximum A Posteriori Estimate, Posterior Predictive Distribution

Approximate Inference

- There are two ways to approximate inference
  - Stochastic method
    - Markov chain Monte Carlo
  - Deterministic method
    - Variational Inference(Variational Bayes) : Approximate the posterior distribution

- Variational Inference : Set the parameter so that Q equals P
  - Define q($\theta|w$) to similar with $P(\theta|D)$
  - KL divergence
  - $D_{KL}(Q||P) = \sum_{x \in X} Q(x) log_b \left(\frac{Q(x)}{P(x)}\right) = \int q(\theta|w) log \frac{q(\theta|w)}{p(\theta|D)} d\theta$

$$= D_{KL} \, [q(\theta|w)||p(\theta)] + E_q[\log p(X)] - E_q[\log p(X|\theta)]$$
$$= D_{KL} \, [q(\theta|w)||p(\theta)] - E_q[\log p(X|\theta)]$$

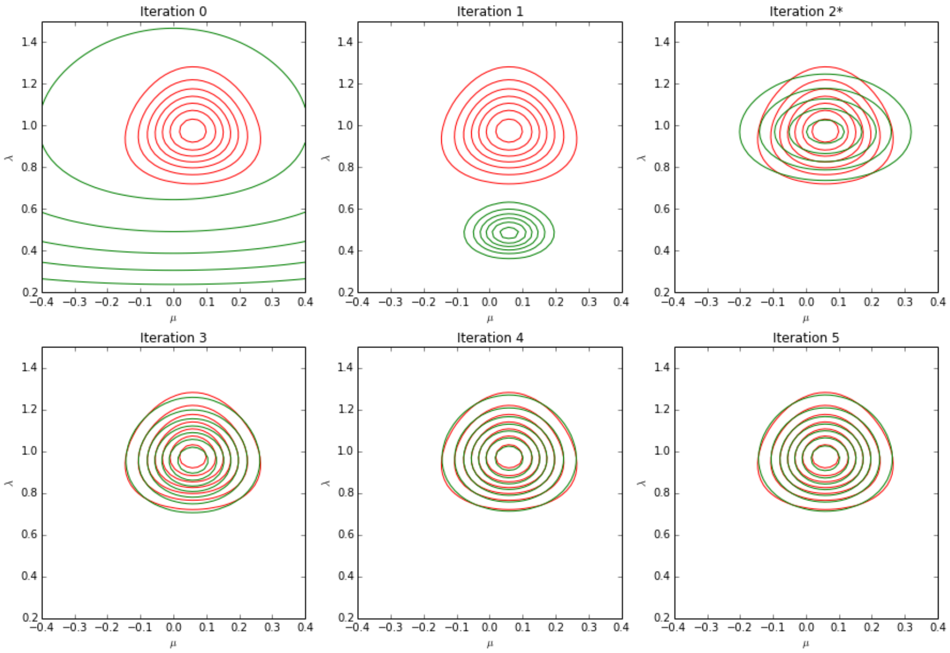
  - $argmin_w(D_{KL}[q(\theta|w)||p(\theta)] - E_q[\log p(X|\theta)])$
  - $[q(\theta|w)||p(\theta)]$ is possible calculating
  - $E_q[\log p(\mathrm{X}|\bar{\theta})]$ is MLE
  - If we find w, we could assume that $q(\theta|w)$ is θ of posterior
  - Learning parameter w

https://hyeongminlee.github.io/post/bnn003_vi/

Variational Inference

- Processing θ_i using variational inference

1. Initialize μ_i, σ_i
2. Parameterize $\theta_i = \mu_i + \sigma_i \times \epsilon_i$
3. Input x, output $\hat{y}$
4. Loss, $L_2 = (y - \hat{y})^2$
5. Backpropagation through $Loss = -\sum_i^N [(1 + \log(\sigma_i{}^2)) - \mu_i{}^2 - \sigma_i{}^2] + (y - \hat{y})^2$
6. Repeat 2~5



https://hyeongminlee.github.io/post/bnn003_vi/
https://github.com/cangermueller/varinf/blob/master/doc/uninorm.ipynb