STAT404 - 베이즈통계입문
Progress Report

평균장 방법과 변분 베이즈

Presenter: 윤민서
데이터과학과 2021320322

Scope and Purpose

- 변분 추론
  - 평균장 방법
  - 변분 베이즈
- 베이지안 최적화

Scope and Purpose

- 목표
  - 수식 전개
  - 두 방법에 대해 각각 서로 다른 subtask에 적용
    - 평균장 방법: 이미지 잡음 제거
    - 변분 베이즈: 선형 회귀
  - 각 방법에서 사용되는 초모수를 베이지안 최적화 방법으로 조정

Literature Review

- 수식 전개
  - https://probml.github.io/pml-book/book0.html

21 Variational inference **731**

21.1 Introduction 731
21.2 Variational inference 732
21.2.1 Alternative interpretations of the variational objective 733
21.2.2 Forward or reverse KL? * 733
21.3 The mean field method 735
21.3.1 Derivation of the mean field update equations 736
21.3.2 Example: mean field for the Ising model 737
21.4 Structured mean field * 739
21.4.1 Example: factorial HMM 740
21.5 Variational Bayes 742
21.5.1 Example: VB for a univariate Gaussian 742
21.5.2 Example: VB for linear regression 746
21.6 Variational Bayes EM 749
21.6.1 Example: VBEM for mixtures of Gaussians * 750
21.7 Variational message passing and VIBES 756
21.8 Local variational bounds * 756
21.8.1 Motivating applications 756
21.8.2 Bohning's quadratic bound to the log-sum-exp function 758
21.8.3 Bounds for the sigmoid function 760
21.8.4 Other bounds and approximations to the log-sum-exp function * 762
21.8.5 Variational inference based on upper bounds 763

Progress

- 평균장 방법

평균장 근사 수식 증명

평균장 근사를 위한 접근에서는 posterior를 fully factorized approximation 형태로 가정한다.

$$q(\mathbf{x}) = \prod_i q_i(\mathbf{x}_i)$$

우리의 목표는 아래의 최적화 문제를 푸는 것이다.

$$\min_{q_1,\dots,q_D} \mathbb{KL}(q||p)$$

즉 각 주변확률분포인 q_i의 모수에 대하여 최적화한다.

변분 추론의 근본적인 목표는 다음의 하계를 최대화하는 것이다.

$$L(q) := -J(q) = \sum_{\mathbf{x}} q(\mathbf{x}) \log \frac{\tilde{p}(\mathbf{x})}{q(\mathbf{x})} \leq \log p(\mathcal{D})$$

이때 q_j를 제외한 나머지 변수를 상수로 취급하면

$$\begin{aligned} L(q_j) &= \sum_{\mathbf{x}} \prod_i q_i(\mathbf{x}_i) \left[\log \tilde{p}(\mathbf{x}) - \sum_k \log q_k(\mathbf{x}_k) \right] \\ &= \sum_{\mathbf{x}_j} \sum_{\mathbf{x}_{-j}} q_j(\mathbf{x}_j) \prod_{i \neq j} q_i(\mathbf{x}_i) \left[\log \tilde{p}(\mathbf{x}) - \sum_k \log q_k(\mathbf{x}_k) \right] \\ &= \sum_{\mathbf{x}_j} q_j(\mathbf{x}_j) \sum_{\mathbf{x}_{-j}} \prod_{i \neq j} q_i(\mathbf{x}_i) \log \tilde{p}(\mathbf{x}) \end{aligned}$$

Progress

- 변분 베이즈

변분 베이즈 수식 증명 - 선형 회귀

Prior: $p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}I)$

Likelihood: $p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{X}\mathbf{w}, \lambda^{-1})$

이때 α와 λ에 대한 추정을 한다고 하면 아래와 같은 prior를 고려할 수 있다.

$$p(\mathbf{w}, \lambda, \alpha) = \mathcal{N}(\mathbf{w}|\mathbf{0}, (\lambda\alpha)^{-1}\mathbf{I})\mathrm{Ga}(\lambda|a_0^{\lambda}, b_0^{\lambda})\mathrm{Ga}(\alpha|a_0^{\alpha}, b_0^{\alpha})$$

또한 posterior q를 계산하기 편한 형태로 인수분해하여 근사하는 방법을 사용할 수 있다.

$$q(\mathbf{w}, \alpha, \lambda) = q(\mathbf{w}, \lambda)q(\alpha)$$

이러한 가정 하에서 Design and Analysis of Learning Classifier Systems의 7장의 내용에 따르면 posterior 아래와 같이 계산할 수 있다.

$$q(\mathbf{w}, \alpha, \lambda) = \mathcal{N}(\mathbf{w}|\mathbf{w}_N, \lambda^{-1}\mathbf{V}_N)\mathrm{Ga}(\lambda|a_N^{\lambda}, b_N^{\lambda})\mathrm{Ga}(\alpha|a_N^{\alpha}, b_N^{\alpha})$$

이때

$$\begin{aligned}
\mathbf{V}_N^{-1} &= \bar{\mathbf{A}} + \mathbf{X}^{\mathbf{X}} \\
\mathbf{w}_N &= \mathbf{V}_N\mathbf{X}^T\mathbf{y} \\
a_N^{\lambda} &= a_0^{\lambda} + \frac{N}{2} \\
b_N^{\lambda} &= b_0^{\lambda} + \frac{1}{2}(||\mathbf{y} - \mathbf{X}\mathbf{w}||^2 + \mathbf{w}_N^T\bar{\mathbf{A}}\mathbf{w}_N) \\
a_N^{\alpha} &= a_0^{\alpha} + \frac{D}{\ }
\end{aligned}$$

Remaining Work

- (저번 Progress report와 비교했을 때) 모델 성능 향상
- 시간이 된다면 더욱 다양한 task에 적용
- 초모수를 베이지안 최적화 방법으로 조정

Expected results

- 베이지안 최적화 방법으로 찾은 초모수가 모델 성능 향상에 상당히 기여를 할 것
  - 격자 탐색, 무작위 탐색에 비하여 시간적으로 효과적
- 사전 분포의 모수도 베이지안 최적화 방법으로 조정할 수 있다면 사전적인 믿음을 부여하는 것을 자동화할 수 있음

Questions?