

Variational Inference

Nipun Batra

November 11, 2023

IIT Gandhinagar

Introduction

Bayesian ML: Recap

- We assume a prior distribution over the parameters of the model given as $P(\theta)$
- We assume a likelihood function $P(D|\theta)$
- We use Bayes' rule to find the posterior distribution of the parameters given the data: $P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$
- Typically, we can not compute the posterior distribution analytically as the denominator is intractable

Bayesian ML: Methods

Laplace Approximation

Approximates the posterior with a Gaussian distribution parameterized by $\Psi = (\mu, \Sigma)$.

$$q_{\Psi}(\theta) = \mathcal{N}(\mu, \Sigma)$$

where μ is the mode of the posterior and Σ is the negative inverse Hessian of the log joint distribution evaluated at θ_{MAP} .

MCMC (Markov Chain Monte Carlo)

Generates samples from the posterior distribution by constructing a Markov chain.

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

Variational Inference

Poses posterior inference as an optimization problem. The approximating distribution is parameterized by Ψ .

$$\Psi^* = \arg \min_{\Psi} \text{KL}(q_{\Psi}(\theta) || P(\theta|D))$$

- KL divergence is a measure of dissimilarity between two distributions.
- It is defined as: $\text{KL}(q||p) = \int q(\theta) \log \frac{q(\theta)}{p(\theta)} d\theta$