MAP estimate on GLMs

DATA :

LINEAR MODEL:

GENERALIZED :
(GLM)

MAXIMUM LIKELIHOOD :
(MLE)

MAP:

$$z_{MAP} = \arg\max_{z} \sum_{n=1}^{N} \log p(y_n / z^T x_n) + \log p(z)$$

Data term

Regularizer

Stochastic Gradient Descent (SGD)

$$\underline{z}_{MAP} = \arg\max_{\underline{z}} \overbrace{\sum_{n=1}^{N} \log p(y_n | \underline{z}^T \underline{x}_n) + \log p(\underline{z})}^{:= L(\underline{z})}$$

Randomly sample an n and compute an "unbiased" estimate

SGD UPDATE: $$z_{t+1} = z_t + \alpha_t \widehat{\frac{\partial L}{\partial z}}\Big|_{z=z_t}$$

Q: converge?

MAP to Bayesian Inference

MAP: $$\underline{z}_{MAP} = \arg\max_{\underline{z}} \sum_{n=1}^{N} \log p(y_n | \underline{z}^T \underline{x}_n) + \log p(\underline{z})$$

Bayes rule: $$p(\underline{z}|\underline{y}) = \frac{p(\underline{y},\underline{z})}{\int p(\underline{y},\underline{z})\, d\underline{z}}$$

Posterior

INTRACTABLE INTEGRAL

Computational issues

GLM: $p(\underline{z}|\underline{y}) = \int \prod_{n=1}^{N} p(y_n / \underline{z}^T \underline{x}_n)\, p(\underline{z})\, d\underline{z}$

POSTERIOR

CONJUGATE PRIORS:

Q: Comp. cost for Gaussian?

NON-CONJUGATE :

WE WILL CONVERT INTEGRATION → OPTIMIZATION

Q: How to approximate?

Question

Why is Bayesian inference computationally challenging?

Variational Inference

MAIN IDEA: $p(\underline{z} \mid \underline{y}) \simeq$ e.g.

LOWER BOUND:

$$\text{VI}: \quad \max_{\lambda} \sum_{n=1}^{N} \mathbb{E}_{q(z;\lambda)} \left[\log p(y_n \mid x_n^T z) \right] - D_{KL} \left[q(z;\lambda) \,\|\, p(z) \right]$$

Data-term Regularizer

Q: Equality?

Question

How does variational inference solve the computational problems with Bayesian Inference?

VI by using SGD

(SEE "VI in five lines in python")

$$\text{MAP:} \quad \max_z \sum_{n=1}^{N} \log p(y_n | z^T x_n) + \log p(z)$$

$$\text{VI:} \quad \max_\lambda \sum_{n=1}^{N} \mathbb{E}_{q(z;\lambda)}\left[\log p(y_n | x_n^T z)\right] - D_{KL}\left[q(z;\lambda) \| p(z)\right]$$

How to compute an unbiased stochastic gradient?

Doubly-stochastic estimate:

$$\text{SGD:} \quad \lambda_{t+1} = \lambda_t + \alpha_t \widehat{\frac{\partial \mathcal{L}}{\partial \lambda}}\Big|_{\lambda=\lambda_t}$$

other methods

- VMP
- SVI

Q: Issues?

Question

Describe one method to perform variational inference.

Discuss computational challenges associated with it.

Mean-Field Inference

ISSUE : size of λ could be > size of z !

EXAMPLE :

SOLUTION :

Mean-field : $q(z;\lambda) = \prod_{i=1}^{D} q_i(z_i;\lambda_i)$

Q: Issues?

Variational Auto-Encoder

Kingma & Welling (ICLR 2014)

Consider unsupervised learning with data $y_n \in \mathbb{R}^D$, $n = 1, \ldots, N$

Variational Auto-Encoder

Variational Lower bound: $\mathbb{E}_{q_\phi}[\log p_\theta(y|z)] - D_{KL}[q_\phi \| p(z)]$

"SGD + variance reduction" works quite well!

SHOW RESULTS

Question

How does variational auto-encoder solve the computational problem with the standard mean-field inference?

Assignment

Question 1: Why is Bayesian inference computationalley challenging?

Question 2: How does variational inference solve the computational problems with Bayesian Inference?

Question 3: Describe one method to perform variational inference. Discuss computational challenges associated with it?

Question 4: How does variational auto-encoder solve the computational problem with the standard mean-field inference?