Laplace approximation

Spring 2024

Instructor: Shandian Zhe
zhe@cs.utah.edu
School of Computing

Outline

- Laplace approximation
- Bayesian logistic regression

Outline

- Laplace approximation
- Bayesian logistic regression

Laplace approximation

- Objective: construct a *Gaussian* distribution to approximate the posterior distribution
- Method: second order Taylor expansion at the posterior mode (i.e., MAP estimation)

Laplace approximation

- Given a joint probability $p(\boldsymbol{\theta}, \mathcal{D})$
- How to compute (approximate) $p(\boldsymbol{\theta}|\mathcal{D})$?

Let us do MAP estimation first

$$\boldsymbol{\theta}_0 = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \ \log p(\boldsymbol{\theta}, \mathcal{D}) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \ \log p(\mathcal{D}|\boldsymbol{\theta})$$

Laplace approximation

- We then expand the log joint probability at the posterior mode

$$f(\boldsymbol{\theta}) \triangleq \log p(\boldsymbol{\theta}, \mathcal{D})$$

$$f(\boldsymbol{\theta}) \approx f(\boldsymbol{\theta}_0) + \boxed{\nabla f(\boldsymbol{\theta}_0)}^\top (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \qquad \nabla f(\boldsymbol{\theta}_0) = \mathbf{0}$$

$$+ \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \boxed{\nabla\nabla f(\boldsymbol{\theta}_0)} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \qquad \nabla\nabla f(\boldsymbol{\theta}_0) \prec 0$$

Why?

$$= f(\boldsymbol{\theta}_0) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{A} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

$$\mathbf{A} = -\nabla\nabla f(\boldsymbol{\theta}_0) \succ 0$$

Laplace approximation

$$f(\boldsymbol{\theta}) \triangleq \log p(\boldsymbol{\theta}, \mathcal{D})$$
$$\approx f(\boldsymbol{\theta}_0) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

$$p(\boldsymbol{\theta}, \mathcal{D}) \approx p(\boldsymbol{\theta}_0, \mathcal{D}) \exp\big(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\big)$$

Gaussian!

$$p(\boldsymbol{\theta}|\mathcal{D}) \approx \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \mathbf{A}^{-1})$$

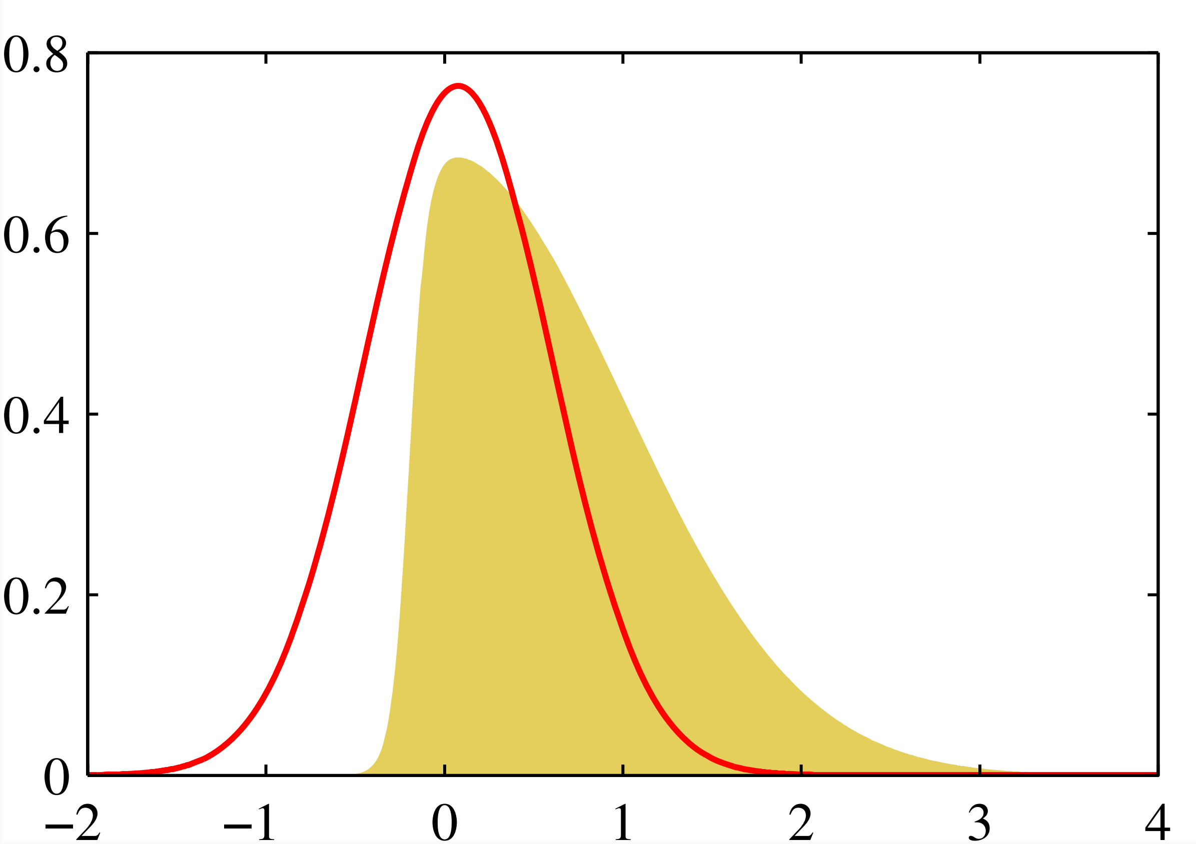
$$\boldsymbol{\theta}_0 = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ \log p(\boldsymbol{\theta}, \mathcal{D})$$
$$\mathbf{A} = -\nabla\nabla \log p(\boldsymbol{\theta}, \mathcal{D})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

Laplace approximation

$$p(z) \propto \exp(-z^2/2)\sigma(20z + 4)$$

Yellow: true

Red: Laplace approx.



Outline

- Laplace approximation
- Bayesian logistic regression

Bayesian Logistic regression

- Given a dataset $\{\phi_n, t_n\}$, where $t_n \in \{0,1\}$, $\phi_n = \phi(\mathbf{x}_n)$ and $n = 1, \ldots, N$, the likelihood function is given by

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0)$$

$$p(\mathbf{t}|\mathbf{w}) = \prod_{n=1}^{N} y_n^{t_n} \{1 - y_n\}^{1-t_n}$$

$$\mathbf{t} = (t_1, \ldots, t_N)^{\mathrm{T}}$$

$$y_n = p(\mathcal{C}_1|\boldsymbol{\phi}_n) = \sigma(\mathbf{w}^\top \boldsymbol{\phi}_n)$$

$$\Downarrow$$

$$p(\mathbf{w}|\mathbf{t}) \propto p(\mathbf{w})p(\mathbf{t}|\mathbf{w})$$

Bayesian logistic regression

$$\log p(\mathbf{w}, \mathbf{t}) \;=\; -\frac{1}{2}(\mathbf{w}-\mathbf{m}_0)^{\mathrm{T}}\mathbf{S}_0^{-1}(\mathbf{w}-\mathbf{m}_0) + \sum_{n=1}^{N}\{t_n \ln y_n + (1-t_n)\ln(1-y_n)\} + \text{const}$$

$$\frac{d\sigma}{da} = \sigma(1-\sigma).$$

$$\mathbf{S}_N = -\nabla\nabla \ln p(\mathbf{w}|\mathbf{t}) = \mathbf{S}_0^{-1} + \sum_{n=1}^{N} y_n(1-y_n)\boldsymbol{\phi}_n\boldsymbol{\phi}_n^{\mathrm{T}}$$

$$q(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{w}_{\mathrm{MAP}}, \mathbf{S}_N^{-1})$$

Bayesian logistic regression

- Predictive distribution: given a new input $\boldsymbol{\phi}$

$$a = \boldsymbol{\phi}^\top \mathbf{w} \qquad\qquad q(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{w}_{\mathrm{MAP}}, \mathbf{S}_N^{-1})$$

$$p(a|\mathbf{t}) \approx \; q(a) = \mathcal{N}(a|\mathbf{w}_{\mathrm{MAP}}^\top \boldsymbol{\phi}, \boldsymbol{\phi}^\top \mathbf{S}_N^{-1} \boldsymbol{\phi})$$

Predictive dist. of score

$$p(\mathcal{C}_1|\mathbf{t}) = \int \sigma(a) q(a) \mathrm{d}a$$

Numerical quadrature

What you need to know

- The general idea of Laplace's Approximation
- Being able to implement it