Math76 Summer 2020

Introduction to Bayesian Computation

Lecture 23:

Hierarchical Modeling

19 August 2020

Hierarchical Inference

Idea: Make model more flexible by including prior and likelihood hyperparameters as additional inference targets. Write Bayes' rule over parameters and "hyperparameters."[1]

↳ additional parameters in prior or likelihood that are often fixed

↳ variance of ε in regression

↳ α, β in a Beta prior

x are standard parameters

θ are the hyperparameters

Standard Model/Inference

$$f(x \mid y, \theta) \propto f(y \mid x, \theta)\, f(x \mid \theta)$$

Hierarchical Model

$$f(x, \theta \mid y) \propto f(y \mid x, \theta)\, f(x, \theta)$$

$$= f(y \mid x, \theta)\, f(x \mid \theta)\, f(\theta)$$

[1] See Gelman et al., "Bayesian Data Analysis" Chapter 5 for much more on hierarchical Bayesian modeling.

Formulation Example – Regression Noise Variance

y = observations of sea ice area (dimension = N)

x = coefficients in regression model

$\theta = \sigma_\varepsilon^2$ Variance of observation noise

$$y = Vx + \varepsilon(\theta) \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2 I)$$

$$\uparrow$$

$$\theta$$

Likelihood

$$f(y|x,\theta) = \left(\frac{1}{(2\pi)^{N/2}|\Sigma_\varepsilon|^{1/2}}\right)\exp\left[-\frac{1}{2}(y-Vx)^T\Sigma_\varepsilon^{-1}(y-Vx)\right]$$

$$\left.\begin{aligned} \Sigma_\varepsilon &= \theta I \\ |\Sigma_\varepsilon| &= \theta^N \\ \Sigma_\varepsilon^{-1} &= \frac{1}{\theta} I \end{aligned}\right\} \Rightarrow f(y|x,\theta) = \frac{1}{(2\pi\theta)^{N/2}}\exp\left[-\frac{1}{2\theta}(y-Vx)^T(y-Vx)\right]$$

Formulation Example – Regression Noise Variance

Prior

"Hyperprior"

$$f(x, \sigma) = f(x|\sigma)\, f(\sigma)$$

$$= f(x)\, f(\sigma)$$

$$f(x) = N(\mu, \Sigma_x) = \frac{1}{(2\pi)^{d/2}|\Sigma_x|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_x)^T \Sigma_x^{-1} (x-\mu_x)\right]$$

Hyperprior

$$f(\sigma) = \text{Inv-Gamma}(\sigma; \alpha, \beta)$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \left(\frac{1}{\sigma}\right)^{\alpha+1} \exp\left[-\frac{\beta}{\sigma}\right]$$

Now have other "hyper-hyper" parameter α, β

Because we can evaluate $f(x, \sigma | y)$ up to a constant

↳ use MCMC to sample posterior

(in this problem, we have a conjugate prior "Normal Inverse Gamma")

Formulation Example – Hospitalization Times

y_i = observed hospitalizations in state i

$g(x_i)$ = model that predicted hospitalizations

x_i = parameters describing hospitalization in state i

- "length-of-stay"
- time to admittance

Noise model

$$g(x_i) + \varepsilon_i = y_i$$

Likelihood

$$\Rightarrow f(y_i \mid x_i)$$

Prior

- only anecdotal observations of hospitalizations

$\bar{\mu}$ = typical hospital stay

$\approx$ 2 weeks

$$f(x_i) = N(\bar{\mu}, \bar{\sigma}^2 I)$$

Observation

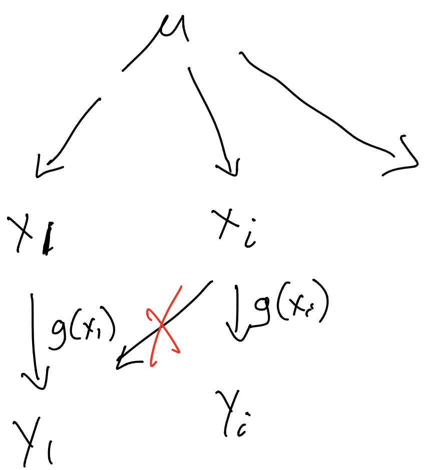
- All NE states should have similar values for x_i
  - can be captured w/ a hierarchical model

Formulation Example – Hospitalization Times

Standard Formulation

$$f(x_i \mid y_i) \propto f(y_i \mid x_i)\, f(x_i)$$

$$f(x_i) = N(\bar{\mu}, \bar{\sigma}^2 I)$$



Layered dependence structure defines "hierarchy" in model

Hierarchical Formulation

- let prior mean vary
  ↳ estimate w/ data from all states

$$f(x_1, x_2, \dots, x_6, \mu \mid y_1, \dots, y_6)$$

$$\propto$$

$$\underbrace{f(y_1, \dots, y_6 \mid x_1, \dots, x_6, \mu)}_{\underbrace{\prod_{i=1}^{6} f(y_i \mid x_i)}_{\text{standard likely}}} \; \underbrace{f(x_1, \dots, x_6, \mu)}_{\prod_{i=1}^{6} f(x_i \mid \mu) f(\mu)}$$

hierarchical prior

Breakout Exercise

1. Ask each person in the group what final project they are working on and how it's going.

2. Discuss possible areas in your projects where you might be able to use a hierarchical model.

3. Nominate one person from the group to describe the possible hierarchical model to the whole class.