10-801: Advanced Topics in Graphcal Models 10-801, Spring 2007

Lecture 2. Hierarchical Bayesian Method

Lecturer: Eric P. Xing *Scribes: Kyung-Ah Sohn*

Example 1 *Hierarchical Bayesian Method*

$$X \sim f(\cdot \mid \theta) = Binomial(n, \theta)$$
$$\theta \sim Beta(\cdot \mid m, m)$$
$$m \sim \pi(m)$$

$$\begin{aligned} P(x) &= \sum_m \int p(x \mid n, \theta) p(\theta \mid m, m) p(m) d\theta \\ &= \sum_m \int \begin{pmatrix} n \\ x \end{pmatrix} \theta^x (1-\theta)^{n-x} \frac{\Gamma(2m)}{\Gamma(m)\Gamma(m)} \theta^{m-1} (1-\theta)^{m-1} \pi(m) d\theta \\ &= \sum_m \int \begin{pmatrix} n \\ x \end{pmatrix} \frac{\Gamma(2m)}{\Gamma(m)\Gamma(m)} \theta^{m+x-1} (1-\theta)^{n+m-x-1} \pi(m) d\theta \\ &= \sum_m \begin{pmatrix} n \\ x \end{pmatrix} \frac{\Gamma(2m)}{\Gamma(m)\Gamma(m)} \pi(m) \end{aligned}$$

Empirical Bayes

The idea is to estimate hyper-parameter $\hat{\theta}$ from data, and apply $\pi(\cdot \mid \hat{\theta})$ to new data and for inference.

Example 2 $X \sim B(X \mid 5, \theta)$
What is the probability $P(0 < \theta_{17} < \epsilon \mid X_{17} = 3)$*? What empirical Bayes does is:*

- *Estimate* α *and* β *from* $P(X_1, \ldots, X_{16} \mid \alpha, \beta)$ *s.t.*

$$\alpha, \beta = argmax\ m(X)$$

- *Then this estimation is used for the posterior inference*

$$P(\theta \mid X) = \frac{P(X \mid \theta) P(\theta \mid \hat{\alpha}, \hat{\beta})}{m(X)}$$

Note: As N goes to infinity, $\delta^{\pi(\hat{\alpha})}(X) \rightarrow \delta^{\pi}(X)$.

Empirical Bayes estimator:

$$\hat{\theta} = \text{argmin}_{\hat{\theta}} \ \text{max}_{\theta} L(\hat{\theta}, \theta)$$

Example 3. James-Stein Estimator

$$X \sim N_p(\theta, I)$$

The least square estimators which minimizes $\sum(\theta - \hat{\theta})^2$ would result in $\hat{\theta}_{LS} = X$. Under the following priors

$$\theta \sim N(0, \tau^2 I)$$
$$\tau^2 \sim \pi$$

the James-Stein estimator is given by

$$\delta_{JS} = \left(1 - \frac{p-2}{\sum X_i^2}\right)\bar{(}X), \quad P \geq 3$$

Example 4.

$$X \sim N(\theta, I)$$
$$\theta \sim N(0, \tau^2 I)$$
$$\Rightarrow \delta_\theta = \left(1 - \frac{1}{1+\tau^2}\right)(\overline{X})$$

where $1 + \tau^2$ is unknown. Here, we can use Empirical Bayes to get this τ.

$$X_i \sim N(0, (1+\tau^2)I)$$

$$T = \frac{\sum_{i=1}^N X_i^2}{1+\tau^2} \sim \chi_p^2 \tag{1}$$

Hence,

$$E(T) = \frac{1}{p-2}$$

We can apply this to (1), which induces:

$$1 + \tau^2 = \frac{p-2}{\sum X_i^2}$$

Admixture model

Admixture model is also known as mixed membership model, or Latent Dirichlet Allocation (LDA).

Mixture model

$$P(X_i \mid Z_i) = N(X_i \mid \mu_{Z_i}, \Sigma_{Z_i})$$
$$Z_i \sim P(\pi_i)$$

$$\begin{aligned} P(X_i) &= \sum_{Z_i} P(X_i \mid Z_i) P(Z_i) \\ &= P(Z_i = 1) N(X_i \mid \mu_1, \Sigma_1) + (1 - P(Z_i = 1))\, N(X_i \mid \mu_2, \Sigma_2) \quad \text{if } Z_i \text{ binary} \end{aligned}$$

Bayesian mixture model

$$\mu \sim N(\mu_0, \alpha_\mu \Sigma^{-1})$$
$$\Sigma^{-1} \sim Wishart(\alpha_T, T) = \frac{1}{e} \exp\left(tr(T\Sigma^{-1}) + \log(\alpha + \mu - 1)\right)$$
$$P(\mu, \Sigma^{-1} \mid X, \ldots) = N(\nu', \Sigma^{-1}) W(\alpha', T')$$

- LDA
- Genetics

These models will be discussed more next time.