Bayesian Non-Parametrics

Advanced Machine Learning for NLP

Jordan Boyd-Graber

OVERVIEW

What about text?

- Gaussian distributions can't model text
- So typically use multinomial distribution as the base distribution
- Remember multinomial:

$$P(N \mid n, \theta) = \frac{n!}{\prod_j N_j!} \prod_j \theta_j^{N_j} \tag{1}$$

What about text?

- Gaussian distributions can't model text [or can they?]
- So typically use multinomial distribution as the base distribution
- Remember multinomial:

$$P(N \mid n, \theta) = \frac{n!}{\prod_j N_j!} \prod_j \theta_j^{N_j} \tag{1}$$

Dirichlet Process Multinomial Mixture Model

- Break off sticks

$$V_1, V_2, \cdots \sim_{\text{iid}} \text{Beta}(1, \alpha) \quad (2)$$

$$C_k \equiv V_k \prod_{j=1}^{k-1} (1 - V_k) \quad (3)$$

Dirichlet Process Multinomial Mixture Model

- Break off sticks

$$V_1, V_2, \cdots \sim_{\text{iid}} \text{Beta}(1, \alpha) \tag{2}$$

$$C_k \equiv V_k \prod_{j=1}^{k-1} (1 - V_k) \tag{3}$$

- Draw atoms

$$\Phi_1, \Phi_2, \cdots \sim_{\text{iid}} \text{Dir}(\beta) \tag{4}$$

Dirichlet Process Multinomial Mixture Model

- Break off sticks

$$V_1, V_2, \cdots \sim_{\text{iid}} \text{Beta}(1, \alpha) \tag{2}$$

$$C_k \equiv V_k \prod_{j=1}^{k-1} (1 - V_k) \tag{3}$$

- Draw atoms

$$\Phi_1, \Phi_2, \cdots \sim_{\text{iid}} \text{Dir}(\beta) \tag{4}$$

- Merge into complete distribution

$$\Theta = \sum_k C_k \delta_{\Phi_k} \tag{5}$$

Dirichlet Process Multinomial Mixture Model

- Break off sticks

$$V_1, V_2, \cdots \sim_{\text{iid}} \text{Beta}(1, \alpha) \tag{2}$$

$$C_k \equiv V_k \prod_{j=1}^{k-1} (1 - V_k) \tag{3}$$

- Draw atoms

$$\Phi_1, \Phi_2, \cdots \sim_{\text{iid}} \text{Dir}(\beta) \tag{4}$$

- Merge into complete distribution

$$\Theta = \sum_k C_k \delta_{\Phi_k} \tag{5}$$

- Draw document word counts

$$\phi_d \sim \Theta \tag{6}$$

$$w_d \sim \phi_d \tag{7}$$

Extending DPMM for text: HDP

- Topic models can use multiple topics per document
- Mixture model can only use one
- HDP is the non-parametric extension

Hierarchical Dirichlet Process

- Draw a global distribution over topics (e.g., $H \equiv \text{Dir}(\alpha)$)

$$G_0 \sim \text{DP}(\gamma, H) \tag{8}$$