

## Note for Graphical Model

Yize Li

- Part 2, Page 7 Belief Propagation (Yi)

The speaker skipped the derivation steps. To understand it, derive it by yourself.

- Part 3, Page 17 Structure learning(Jessica & Xing)

It is intractable to try every candidate structure. So we need make some assumption on the general structure and then perform greedy search.

- Part 4, Page 3 Laplace Approximation (Yi)

Laplace Approximation can be used for model comparison /selection. Here, 'model' could be the number of parameters, relationship between nodes. This method can guide feature selection process.

- Part 4, Page 9 (Sarah)

Focusing on one term,  $q_j$ , we can write this as:

$$\mathcal{F}(q_j) = \int q_j(H_j) \langle \log p(H, V) \rangle_{\sim q_j(H_j)} dH_j + \int q_j(H_j) \log q_j(H_j) dH_j + \text{const}$$

Where  $\langle \cdot \rangle_{\sim q_j(H_j)}$  denotes averaging w.r.t.  $q_i(H_i)$  for all  $i \neq j$

The first term is calculating the integral for  $q_j(H_j)$ . This is a tractable solution to focus on one term while keeping others fixed.

- Part4, Page 14 Results comparison (Yi)

Models' performance may change with the training data size. A lot of conclusion is data set dependent. When we conduct experiments, we should be very careful about data.

- Demo (Yi)

The speaker showed a demo for model selection. It is a good example about how to debug our model.

- Directed Graphical Model vs. Social network Model (Jian)

The nodes in directed graphical model are random variables. Each node may be mapped to several instants in dataset. In social network model, each node denotes an instant (e.g a person).

One of the important properties of graphical model is conditional independence. We can utilize this property in personalized information retrieval. For example, for each user, we can collect a lot of information, however, that would be every costly. Instead, we only use the information from the user's community (Here, you can consider the community as user's Markov Boundry)

- Dirichlet Prior (Yize)

Dirichlet is popular: (1) it has some good properties( conjugate to the multinomial distribution); (2)Dirichlet prior + Multinomial Model works well in IR. If features are "bag of words", Dirichlet prior is a good choice.

- Hidden variables in Graphical Model (Lanbo)

Generally, hidden variables are variables that are not directly observed, such as hidden parameters (eg, hidden representation for users and items in collaborative filtering), prior, and missing value in the data.

- Undirected graph is easier to learn than directed graph(refer to deep learning) due to Markov Blanket property