Machine Learning for Data Science (CS4786)
Lecture 27

Last Lecture

Course Webpage :
http://www.cs.cornell.edu/Courses/cs4786/2016sp/

Announcements

- Competition II deadline is a hard one, no extensions!
- Keep report to 5 pages or so
- Big hint: use HMMs

- Assignment 3, message passing question has been made optional (bonus).
- Points scored here can be used to compensate for points lost in other assignments or competitions

Message Passing

- Yes, we normalize our messages, but this is not always required as you might have found in other sources.
- You can use either of $m_{1\to 2}$ or λ_{X_i}, π_{X_i} notations for assignment, we did both of these in class.

Belief Propagation

1. For every observation $X_j = x_j$ define $E_{X_j}(x) = \mathbf{1}\left\{x = x_j\right\}$, for unobserved variables set $E_{X_j}(x) = 1$
2. At round 0, all messages between nodes are 1

Belief Propagation

Message to Parent X_j

$$\lambda_{X_i}(u_j) \propto \sum_{x} \sum_{u \setminus u_j} E_{X_i}(x) P(X_i = x | \text{Parent}(X_i) = u) \left(\prod_{k \in \text{children}} \lambda_{X_k}(x) \prod_{k \in \text{Parent}(X_i), k \neq j} \pi_{X_i}(u_k) \right)$$

$$\sum_{x, \text{all parents but } \mathrm{X_j}} E_{X_i}(x) P(X_i = x | \text{Parent}(X_i) = u) (\text{product of all messages but one from } \mathrm{X_j})$$

Message to child X_j

$$\pi_{X_i}(x) \propto E_{X_i}(x) \sum_{u} P(X_i = x | \text{Parent}(X_i) = u) \left(\prod_{k \in \text{Parent}(X_i)} \pi_{X_i}(u_k) \prod_{k \in \text{children}, k \neq j} \lambda_{X_k}(x) \right)$$

$$\sum_{\text{all parents}} E_{X_i}(x) P(X_i = x | \text{Parent}(X_i) = u) (\text{product of all messages but one from } \mathrm{X_j})$$

Example

Lessons Learnt

Real World Data

1. Real world data can be hard, Competition I!
2. Good features is key! Requires domain knowledge.
3. Right tool for the right job!
4. Understanding your data is crucial!

No Free Lunch

- No model is universally good or better
- To make good models we need to make good asssumptions
- Examples:
  - Probabilistic model generating the data
  - On relationship between various variables
  - Use the right latent variables in the model

Tools You Have

- Probabilistic modeling of data
- EM algorithm (power of wishful thinking)
- Graphical models (to model intuitive relationship between variables)
- Algorithms for dimensionality reduction: to reduce reconstruction error, to retain common information bettween multiple sources
- Algorithm for clustering nodes of a graph (spectral clustering), for extracting features for nodes in a graph (spectral embedding)

Thanks!