Lecture 17: Loopy BP, Gibbs Sampling, and VI with Gradients

Lecturer: Sasha Rush   *Scribes: Patrick Varin*

17.1 Loopy BP

17.1.1 History

The history of Loopy-BP begain in 1988 with Judea Pearl who tried to analyze the behavior of the BP algorithm (which gives exact marginal inference on tree) on graphs that are not trees, like the Ising model. Remember the message passing algorithms is

$$m_{s \to t}(x_t) = \sum_{x_s} \psi(x_s)\psi(x_s, x_t) \prod_{u \in \text{NBR}(s)-t} m_{u \to s}(x_s)$$

$$bel_s(x_s) \propto \psi(x_s) \prod_{t \in \text{NBR}(s)} m_{t \to s}(x_s)$$

Note that this algorithm doesn't require an ordering on the nodes, so it can naturally extend to cyclic graphs. Around 1998, a decade after Pearl raised the question of the BP algorithm on general graphs, papers studying *Turbo coding* or *Low density parity check* (LDPC) codes used the BP algorithm on cyclical graphs with empirical success, motivating more research. This research drew parallels between loopy-BP and variational inference.

17.1.2 Implementations

We can implement loopy belief propagation either *synchronously* or *asynchronously*

- Synchronous Updates: All of the nodes are updated togethers, so that every node at iteration t depends on it's Markov blanket at $t-1$

  – parallelizable
  – usually takes more updates

- Asynchronous Updates: The nodes are given an order and updated sequentially

  – sequential
  – usually converges in fewer updates

These updates look similar to mean field variational inference with some differences in performance.

- Loopy BP is exact on trees, mean-field variational inference is not.

- Loopy BP is not guaranteed to converge, where VI will.

- If Loopy BP converges it is not guaranteed to converge to a local optimum.

Loopy BP is generally more costly than mean field VI, because it stores node beliefs in addition to edge messages, whereas mean field VI only stores node parameters. If the graph is large and dense, then Loopy BP might be very costly.
In practice, because Loopy BP contains more information about the problem than mean field VI it tends to work better.

17.2   Gibbs Sampling (Preview)

17.3   Variational Inference with Gradients