Particle methods

CS B553
Spring 2013

Announcements

- A3 posted
  - Due Friday March 8, 11:59PM

Assignment 3: Unary potentials

- The skeleton code automatically computes the unary potentials for you
  - You don’t have to know the details, but it might be helpful to have a general idea
- The code loads in the training images and ground truth part locations
  - Then it learns little templates of what the 6 parts “look like”
  - Each pixel of template has a mean and variance; e.g. means:

Unary potentials

\[ \psi_1(L_1, I) \]

Unary potentials

\[ \psi_2(L_2, I) \]
Unary potentials

\(\Psi(3, i)\)

Unary potentials

\(\Psi(4, i)\)

Unary potentials

\(\Psi(5, i)\)

Unary potentials

\(\Psi(6, i)\)

Naïve result

Making inference tractable

• In practice, making inference tractable is a key challenge in applying graphical models to applications
• Typically, the options are:
  – Exact inference with arbitrary potentials on a graphical model, but with a simplified structure
  – Exact inference on a graphical model with arbitrary structure, but restricted potentials
  – Graphical model with arbitrary structure and arbitrary potentials, but settle for approximate inference
Particle-based techniques

• A particle is an assignment of values to (some) variables of a graphical model
  – Full particles: assignments of values to all variables
  – Collapsed particles: assignments to some variables

• Basic idea: Sets of particles can be used to approximate a distribution
  – E.g. Many samples from a distribution can be a good representation of original distribution

Forward sampling

• For a Bayes net, we can sample particles using the simple Forward sampling algorithm
  – Sample values from priors at root nodes
  – For a node X for which values have been sampled for all parents, sample from P(X | Parents(X))

Computing marginals

• Forward sampling gives a very simple technique for computing marginals over set of variables X
  – Collect many marginals using Forward sampling
  – For each possible value of X, count the percentage of sampled particles that have that value

Example (from A2)

Based on these samples, we can approximate:
  – P(S1=PRON) = 1
  – P(S3=PRT)=0.2, P(S3=ADP)=0.8
  – P(S3=ADP, S10=VERB) = 0.2, P(S3=ADP, S10=NOUN) = 0.6

Approximation error

• Clearly the approximation error will decrease as number of particles increases
  – What is the precise relationship?

Approximation bounds

Consider the cases where the approximation error is below the desired bound (\(\varepsilon\)). The table shows the number of samples needed for different levels of certainty.

<table>
<thead>
<tr>
<th>Desired bound ((\varepsilon))</th>
<th>99% certainty</th>
<th>95% certainty</th>
<th>80% certainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>10</td>
<td>40</td>
<td>100</td>
</tr>
<tr>
<td>0.01</td>
<td>20</td>
<td>80</td>
<td>200</td>
</tr>
<tr>
<td>0.005</td>
<td>40</td>
<td>160</td>
<td>400</td>
</tr>
</tbody>
</table>
Handling evidence

• In general, we’re interested in computing marginals conditioned on some evidence, i.e. $P(X | Y=y)$
• One easy way to do this with forward sampling:
  – Sample many particles from the Bayes net
  – If a particle has $Y=y$, then keep it; else discard it
  – Compute marginals as before, using only the remaining particles
• Disadvantages of this approach?

Handling evidence more efficiently

• Say we observe $SAT=s1$
• Obvious idea:
  – When we reach an observed variable, simply set it to observed value without sampling
• What’s the problem with this?

Likelihood weighting

• Compute a likelihood weight for each particle
  – Initial weight=1
  – Sample values from priors at root nodes
  – For unobserved $X$ for which values have been sampled for all parents, sample from $P(X | \text{Parents}(X))$
  – For observed $Y=y$, set $Y=y$ but then update weight: $w = w \times P(Y=y | \text{Parents}(Y))$

Computing marginals with LW

• LW produces a weighted set of particles
  – To compute $P(X=x | Y=y)$, take sum of weights of particles with $X=x$, over sum of weights of all sampled particles
• Given samples N samples $(\xi_1, w_1), (\xi_2, w_2), \ldots, (\xi_N, w_N)$,
  \[
  \hat{P}(X=x | Y=y) = \frac{\sum_{i=1}^{N} w_i I(x_i = x)}{\sum_{i=1}^{N} w_i}
  \]
  – where $x_i$ refers to the $x$ variables of sample $\xi_i$, and $I$ is an indicator function that is 1 if the two sets of values are equal