Particle methods

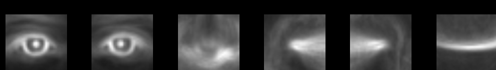
CS B553
Spring 2013

Announcements

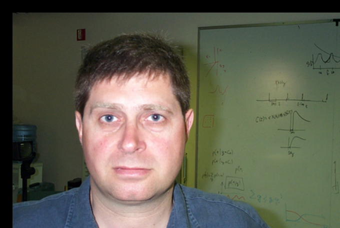
- A3 posted
  - Due Friday March 8, 11:59PM

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



Unary potentials

Ψ1(L1, I)

Unary potentials

Ψ2(L2, I)

Unary potentials

$\Psi 3(L3, I)$

Unary potentials

$\Psi 4(L4, I)$

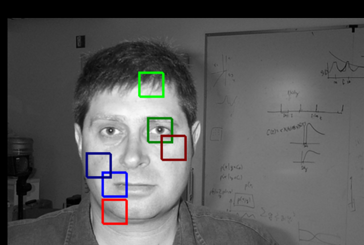
Unary potentials

$\Psi 5(L5, I)$

Unary potentials

$\Psi 6(L6, I)$

Naïve result



Making inference tractable

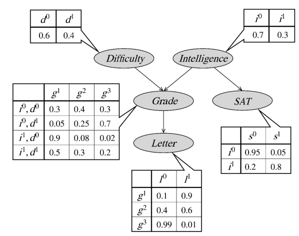
- In practice, making inference tractable is a key challenge in applying graphical models to applications
- Typically, the options are:
  - **Exact inference** with **arbitrary potentials** on a graphical model, but with a **simplified structure**
  - **Exact inference** on a graphical model with **arbitrary 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 particles using Forward sampling
  - For each possible value of X, count the percentage of sampled particles that have that value

Example (from A2)

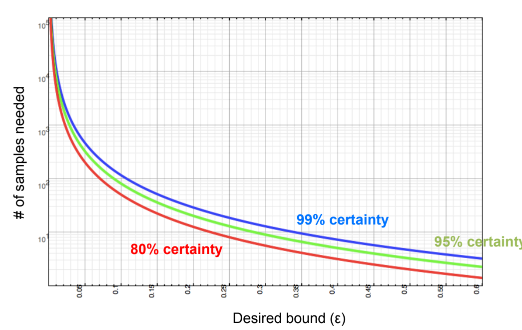
```
--------------
Considering sentence: She started to brush the dirt and bits of leaves off her clothes .
Ground truth: PRON VERB PRT  VERB DET  NOUN CONJ NOUN ADP  NOUN ADP  DET  NOUN .
Naive:        PRON VERB PRT  NOUN DET  NOUN CONJ NOUN ADP  VERB PRT  DET  NOUN .
Bayes:        PRON VERB PRT  VERB DET  NOUN CONJ NOUN ADP  NOUN ADP  DET  NOUN .
Sample 1:     PRON VERB PRT  VERB DET  NOUN CONJ NOUN ADP  NOUN ADP  DET  NOUN .
Sample 2:     PRON VERB ADP  NOUN DET  NOUN CONJ NOUN ADP  NOUN ADP  DET  NOUN .
Sample 3:     PRON VERB ADP  NOUN DET  NOUN CONJ NOUN ADP  VERB PRT  DET  NOUN .
Sample 4:     PRON VERB ADP  NOUN DET  NOUN CONJ NOUN ADP  NOUN ADP  DET  NOUN .
Sample 5:     PRON VERB ADP  NOUN DET  NOUN CONJ NOUN ADP  NOUN ADP  DET  NOUN .
```

- 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

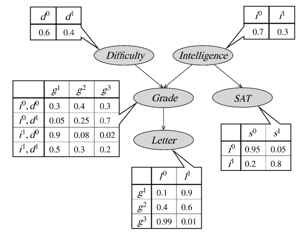


Handling evidence

- 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?

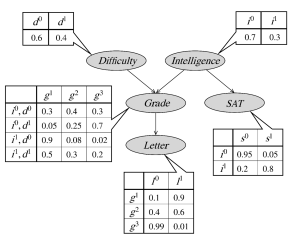
Handling evidence more efficiently

- 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 | Parents(X))
  - For observed Y=y, set Y=y but then update weight: w=w * P(Y=y | 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 $(\xi_1, w_1), (\xi_2, w_2), ..., (\xi_N, x_N)$,

$$\hat{P}(\mathbf{X}=\mathbf{x}|\mathbf{Y}=\mathbf{y}) = \frac{\sum_{i=1}^{N} w[i] \;\; I(\mathbf{x}[i]=\mathbf{x})}{\sum_{i=1}^{N} w[i]}$$

  - where $\mathbf{x}[i]$ 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