HW 3 Review
Bayesian Models

Computational Cognitive Modeling

April 7th 2020

HW3 Overview

A. Bayesian Concept Learning (Number Game) – 80 pts

B. Probabilistic Programs – 40 pts

C. Metropolis Hastings & speech perception – 30 pts

Part A

Part A: The Number Game

1 random “yes” example

60

- An unknown computer programs generates numbers in 1 to 100
- We’re provided with a small set of random examples
- Which other numbers will be accepted by the program?

Inference regarding hypothesized concepts (h):

Posterior:

$$P(h|X) = \frac{P(X|h)P(h)}{\sum_{h' \in H} P(X|h')P(h')}$$

Likelihood:

$$P(X|h) = \prod_{i=1}^{n} P(x^{(i)}|h).$$

$$P(x^{(i)}|h) = \frac{1}{|h|} \text{ if } x^{(i)} \in h \quad \text{(else zero)}$$

Problems 1 & 2: Show your work

1. Compute the posterior probability distribution by hand
   - We know the program accepts X = {10, 30}

2. Make posterior predictions about new numbers y

$$P(y \in C \mid X) = \sum_{h \in H} P(y \in C \mid h) P(h|X),$$

Posterior probability of h

Problems 3 onward: Implementation

Things that are useful to understand:

- How the hypotheses are initialized & stored
- The use of the lambda parameter
- convert_h_list_to_numpy
- Computing probabilities in log space and using logsumexp

*****Let's look at the code*****

Problems 7: Likelihood weighted sampling

- Sample hypotheses $h^{(m)}$ from a surrogate distribution *Q(h)*
- Re-weight the samples to approximate your target posterior *P(h|D).*

$$E[\phi(h)|D] \approx \frac{1}{\sum_m w^{(m)}} \sum_m w^{(m)} \phi(h^{(m)})$$

where samples $h^{(1)}, \ldots, h^{(M)}$ are generated from $Q(h^{(m)})$

We can set Q to be the prior, in order to get **"Likelihood weighted sampling"**

Example: likelihood weighted sampling

16

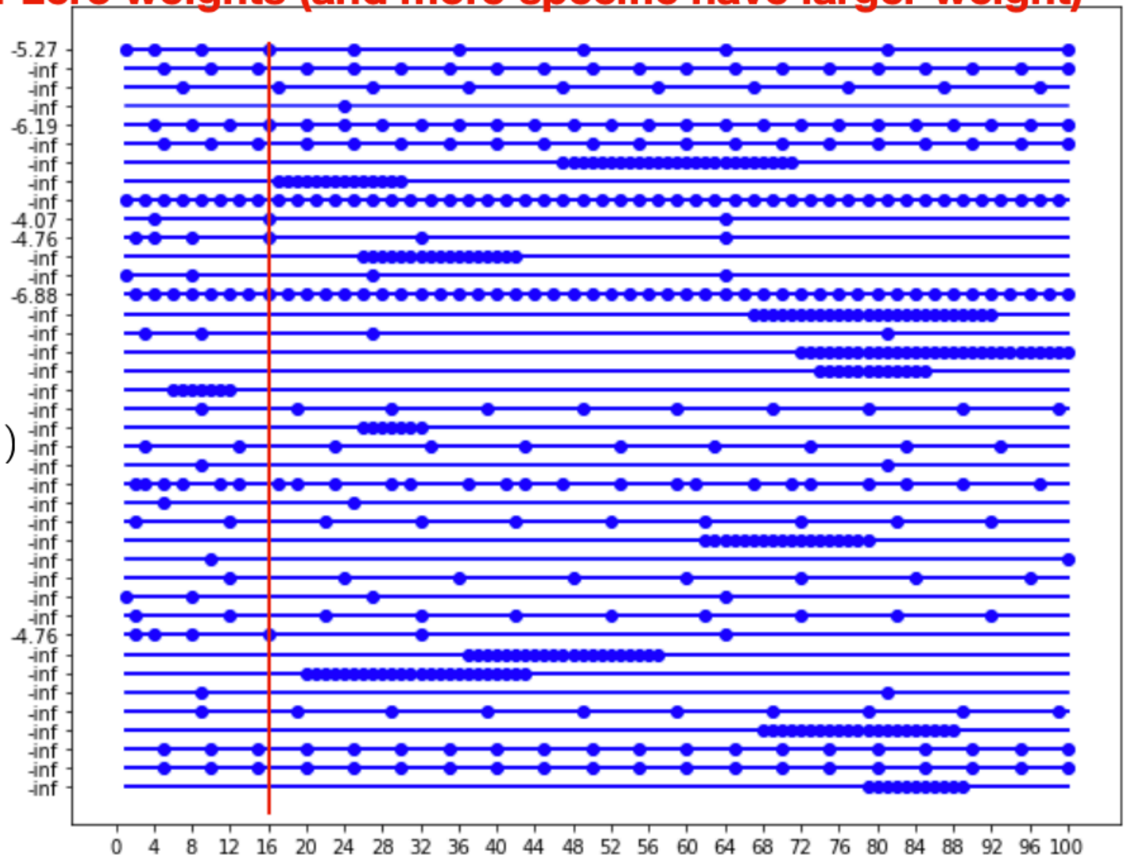
rows are samples from prior

$h^{(m)}$

$$Q(h) \leftarrow P(h)$$

non-zero weights (and more specific have larger weight)

$\log(w^{(m)})$

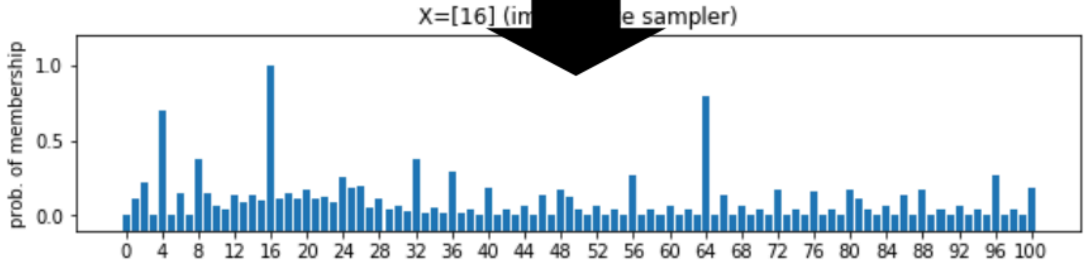


$$E[\phi(h)|D] \approx \frac{1}{\sum_m w^{(m)}} \sum_m w^{(m)} \phi(h^{(m)})$$

$$\phi(h) = 1\{y \in C\}$$

is new number y in the hypothesis?

compute prob. of membership (weighted average)



Problem 7:

i. Draw samples of h based on prior probabilities

ii. Calculate the appropriate weight for each sample

iii. Compute the weighted sum of samples for predicting of y is in the concept

Part B

Part B: Probabilistic programs

- A series of tennis matches are described in a probabilistic model
- We want to use Bayesian inference to reason about parameters in this model, such as a player’s strength

Design of the model:

- Every player has a strength value sampled from Gaussian
- Team strength is the sum of two players’ strengths
- Random laziness cuts a player’s strength in half
- Winners: team with the greater strength

look at the code

Part B: Rejection sampling

- Sample hypotheses from the prior P(h) and data $D^{(m)}$ from the likelihood P(D|h)
- If your sample data $D^{(m)}$ match the target (observed) data, store $h^{(m)}$ as a sample from the posterior

Simple algorithm for a rejection sampler:

$m \leftarrow 1$
while $m < M$ **do**
  sample $h^{(m)} \sim P(h)$
  sample $D^{(m)} \sim P(D|h^{(m)})$
  if $D^{(m)}$ and D match exactly **then** *(if sampled and real data match)*
    accept $h^{(m)}$ as a sample
    $m \leftarrow m + 1$
  end if
end while

What does this look like in the context of the tennis problem?

- We can sample a state of the world (player strengths, laziness, match results, etc.)
- *What are $D^{(m)}$ and D?*

Part C

Part C: Metropolis Hastings

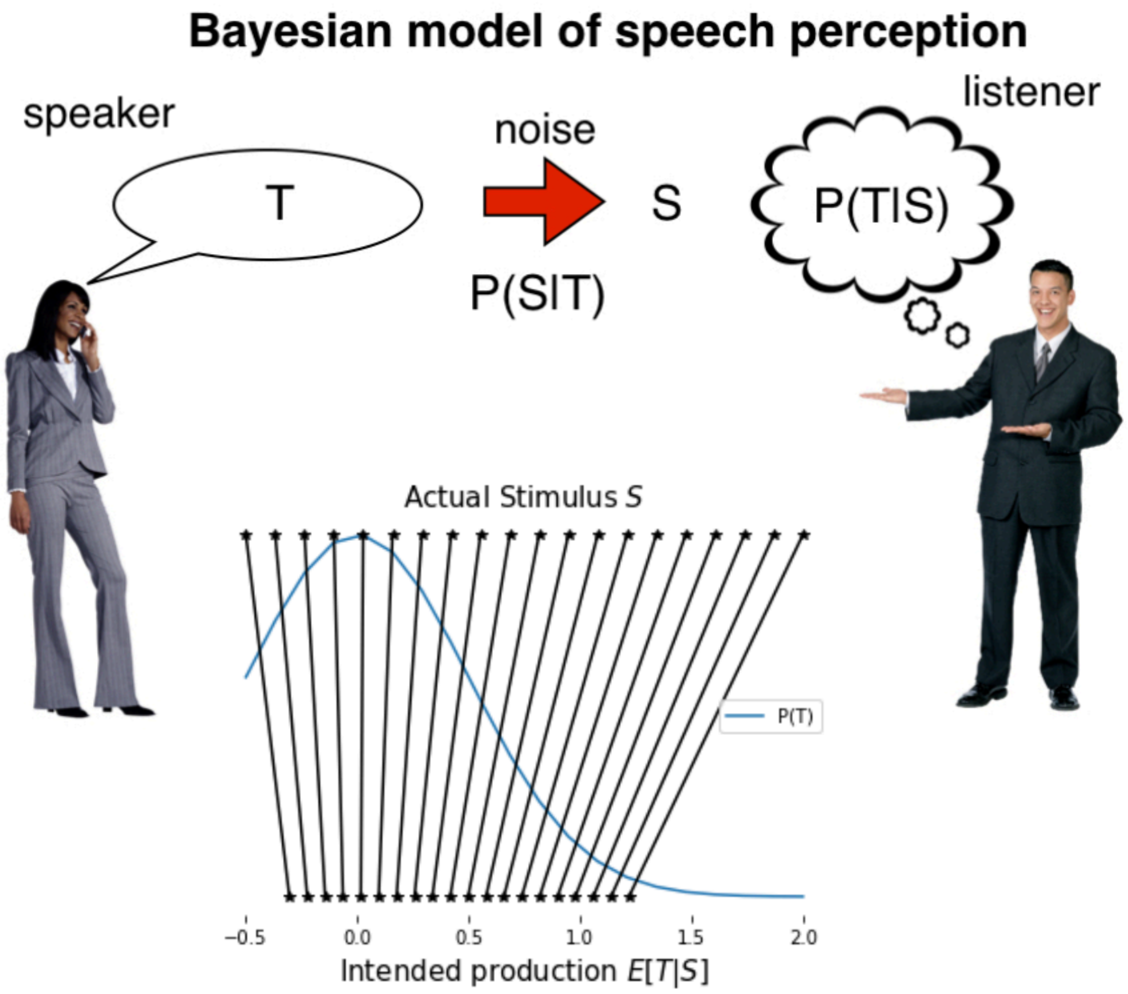
- A speaker produces sound T
- Due to noise, listener receives S

A Bayesian model of speech perception:
- A listener's perceptual system estimates P(T|S)

We could model this with exact Bayesian inference:
- Prior and Likelihood both Gaussian

$$P(T|S) = \frac{P(S|T)P(T)}{P(S)}. \qquad P(T|S) = N\left(\frac{\sigma_c^2 S + \sigma_S^2 \mu_c}{\sigma_c^2 + \sigma_S^2}, \frac{\sigma_c^2 \sigma_S^2}{\sigma_c^2 + \sigma_S^2}\right).$$

But in this problem, we use Metropolis Hastings



Part C: Metropolis Hastings

Full Metropolis-Hastings algorithm:

pick initial $h^{(1)}$
for $t \leftarrow 1 \dots (T-1)$ **do**
  sample $h' \sim Q(h'|h^{(t)})$
  $a = \frac{P(h'|D)Q(h^{(t)}|h')}{P(h^{(t)}|D)Q(h'|h^{(t)})}$
  if $a \geq 1$ **then**
    $h^{(t+1)} \leftarrow h'$
  else
    $h^{(t+1)} \leftarrow h'$ with probability a
    otherwise, $h^{(t+1)} \leftarrow h^{(t)}$
  end if
end for

In the speech perception problem:

- Initial h = an initial value from X
- h' is sampled from a proposal distribution:
  - Gaussian (mean = h, sd = 0.25)
- We can calculate the posterior probabilities of h and h' using what we have (prior & likelihood)
- And then calculate the acceptance ratio
- Decide to keep h or h', and repeat

Acceptance ratio:

For our purposes:

$$a = \frac{P(h'|D)}{P(h^{(t)}|D)}$$

Questions?