Inference And Sampling

Arnav Gupta

November 27, 2024

Contents

1 Hidden Markov Models

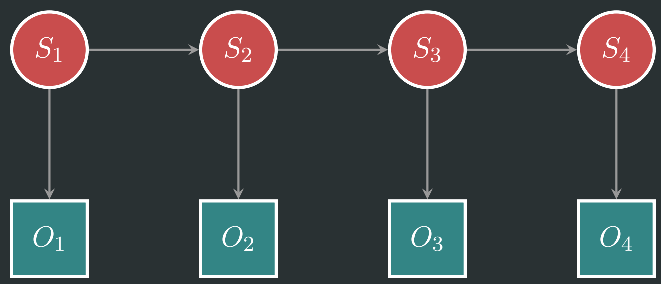
A node may repeat over time, so this requires an explicit encoding of time. A chain has a length equal to the amount of time modeled. Times can be **event-driven** or **clock-driven**.

The **Markov assumption**

$$P(S_{t+1} \mid S_1, \ldots, S_t) = P(S_{t+1} \mid S_t)$$

gives the dynamics of the Markov chain.

In a **Hidden Markov Model (HMM)** observations O_t and observation functions $P(O_t \mid S_t)$ are added to the states. The observations are always observed, so their nodes are square.



Given a sequence of observations $O_1, \ldots, O_t$, the following can be estimated:

$$P(S_t \mid O_1, \ldots, O_t)$$

and

$$P(S_k \mid O_1, \ldots, O_k) \text{ for } k < t$$

The most well-known application of HMMs is in speech recognition. These can be built in a hierarchical model, with higher-level models having words as states to built sentences as HMMs and lower-level models having phonemes as states to have words as HMMs. Observations are generally audio features, with different levels of granularity of audio features being present in different levels of models.

1.1 Belief Monitoring

1.1.1 Forward

$$\alpha_t(i) = P(O_{1\ldots t}, S_t = i) = P(O_t \mid S_t = i) \sum_{S_{t-1}} P(S_t = i \mid s_{t-1}) \alpha_{t-1}(s_{t-1})$$

1.1.2 Backward

$$\beta_t(i) = P(O_{t+1\ldots T} \mid S_t = i) = \sum_{s_{t+1}} \beta_{t+1}(s_{t+1}) P(O_{t+1} \mid s_{t+1}) P(s_{t+1} \mid S_t = i)$$

1.1.3 Forward-Backward

$$\alpha_t(i)\beta_t(i) = P(O_{1\ldots T}, S_t = i) \propto P(S_t = i \mid O)$$

2 Dynamic Bayesian Networks

Any Bayesian network can repeat over time, which is a **Dynamic Bayesian Network**. Many examples can be solved with variable elimination though they can become too complex with enough variables.

Bayesian probability ensures that evidence is integrated proportionally to its precision, so sensors are **precision weighted**.

Variable elimination is an exact algorithm, and sometimes these calculations can be difficult or probability distributions are unknown. This requires resorting to stochastic sampling, since sampling from a distribution allows for estimation.

3 Sampling

With **stochastic simulation**, get probabilities from samples. Specifically, sample from a variable's posterior probability to estimate its posterior probability.

To generate samples from a distribution for a variable X with a discrete domain or a real domain:

- totally order values of the domain of X
- generate the cumulative probability distribution $f(x) = P(X \leq x)$
- select a value y uniformly in the range $[0, 1]$
- select the x such that $f(x) = y$

3.1 Forward Sampling

Sample variables one at a time, specifically sampling parents before sampling a node. Given values for the parents of X, sample from the probability of X given its parents.

For samples s_i, $i = 1 \ldots N$

$$P(X = x_i) \propto \sum_{s_i} \delta(x_i) = N_{X=x_i}$$

where $\delta(x_i)$ is 1 if $X = x_i$ in s_i and 0 otherwise.

Inference via sampling approaches the probability as the number of samples increases.