Announcements

Assignment V posted (finally).

Sampling based inference

- Resources.
  - Bishop, chapter 11
  - Kollar and Friedman, chapter 12
  - Andrieu et al. (linked to on lecture page).

- Kollar and Friedman uses “particles” terminology instead of “samples”.

Sampling based inference

- We have studied two themes in inference.
  - Marginalization / expectation / summing out or integration
  - Optimization

- Two flavors of activities
  - Fitting (inference using a model)
  - Learning (inference to find a model)

- These activities are basically the same in the generative modeling approach.

Motivation for sampling methods

- Real problems are typically complex and high dimensional.

- Example, images as evidence for stuff in the world
Motivation for sampling methods

• Real problems are typically complex and high dimensional.

• Suppose that we could generate samples from a distribution that is proportional to one we are interested in.
  Typical case we are often interested in is $p(\theta|D)$

$$p(\theta|D) = \frac{p(\theta)p(D|\theta)}{p(D)}$$

Consider $\tilde{p}(z) = p(\theta)p(D|\theta)$

Motivation for sampling methods (II)

• Now consider computing the expectation of a function $f(z)$ over $p(z)$.

• Recall that this looks like

$$E_{p(z)}[f] = \int_z f(z)p(z)dz$$

• How can we approximate or estimate $E$?

Motivation for sampling methods (II)

• Now consider computing the expectation of a function $f(z)$ over $p(z)$.

• Recall that this looks like

$$E_{p(z)}[f] = \int_z f(z)p(z)dz$$

• A bad plan for computing $E$:

  Discretize the space where $z$ lives into $L$ blocks

  Then compute

  $$E_{p(z)}[f] \equiv \frac{1}{L} \sum_{i=1}^L p(z)f(z)$$
Motivation for sampling methods (II)

- Now consider computing the expectation of a function \( f(z) \) over \( p(z) \).
- Recall that this looks like
  \[
  E_{p(z)}[f] = \int f(z)p(z)\,dz
  \]
- A better plan, assuming we can sample \( \tilde{p}(z) \)
  
  Given independent samples \( z^{(i)} \) from \( \tilde{p}(z) \)
  
  Estimate
  \[
  E_{\tilde{p}(z)}[f] \approx \frac{1}{L} \sum_{i=1}^{L} f(z)
  \]

Challenges for sampling

In real problems sampling \( p(z) \) is very difficult.

We typically do not know the normalization constant, \( Z \).
(So we need to use \( \tilde{p}(z) \)).

Even if we can draw samples, it is hard to know if (when) they are good, and if we have enough of them.

Evaluating \( \tilde{p}(z) \) is generally much easier (although, it can also be quite involved).

Sampling framework

We assume that sampling from \( \tilde{p}(z) \) is hard, but that evaluating \( \tilde{p}(z) \) is relatively easy.

We also assume that the dimension of \( z \) is high, and that \( \tilde{p}(z) \) may not have closed from (but we can evaluate it).

We will develop the material in the context of computing expectations, but sampling also supports picking a good answer, such as a MAP estimate of parameters.

Basic Sampling (so far)

- Uniform sampling (everything builds on this)
- Sampling from a multinomial
- Sampling for selected other distributions (e.g., Gaussian)
  - At least, Matlab knows how to do it.
- Sampling univariate distributions using the inverse of the cumulative distribution.
Basic Sampling (so far)

• Sampling univariate distributions using the inverse of the cumulative distribution.

\[ p(y) \quad h(y) \]
\[ y \quad 0 \quad 1 \]

Basic Sampling (so far)

• Sampling directed graphical models using ancestral sampling.

Rejection Sampling

Assume that we have an easy to sample function, , and a constant, \( k \), where we know that \( p(z) \leq k \cdot q(z) \).

1) Sample \( q(z) \)
2) Keep samples in proportion to \( \frac{p(z)}{k \cdot q(z)} \) and reject the rest.

Rejection Sampling

1) Sample \( q(z) \)
2) Keep samples in proportion to \( \frac{p(z)}{k \cdot q(z)} \) and reject the rest.
Rejection Sampling

- Rejection sampling is hopeless in high dimensions, but is useful for sampling low dimensional “building block” functions.
- E.G., the Box-Muller method for generating samples from a Gaussian uses rejection sampling.

A second example where a gamma distribution is approximated by a Cauchy proposal distribution.

Importance Sampling

Rewrite $E_{p \leftarrow q}[f] = \int f(z) p(z) dz$

$= \int f(z) \frac{p(z)}{q(z)} q(z) dz$

$= \frac{1}{L} \sum_{i=1}^{L} \frac{p(z^{(i)})}{q(z^{(i)})} f(z^{(i)})$ where samples come from $q(z)$

For complex functions, a good $q()$ and $k$ may not be available.
- One attempt to adaptively find a good $q()$ (see Bishop 11.1.3)