

successive samples.

17.4 Gibbs Sampling

So far we have described how to draw samples from a distribution $q(\mathbf{x})$ by repeatedly updating $\mathbf{x} \leftarrow \mathbf{x}' \sim T(\mathbf{x}' | \mathbf{x})$. However, we have not described how to ensure that $q(\mathbf{x})$ is a useful distribution. Two basic approaches are considered in this book. The first one is to derive T from a given learned p_{model} , described below with the case of sampling from EBMs. The second one is to directly parametrize T and learn it, so that its stationary distribution implicitly defines the p_{model} of interest. Examples of this second approach are discussed in sections 20.12 and 20.13.

In the context of deep learning, we commonly use Markov chains to draw samples from an energy-based model defining a distribution $p_{\text{model}}(\mathbf{x})$. In this case, we want the $q(\mathbf{x})$ for the Markov chain to be $p_{\text{model}}(\mathbf{x})$. To obtain the desired $q(\mathbf{x})$, we must choose an appropriate $T(\mathbf{x}' | \mathbf{x})$.

A conceptually simple and effective approach to building a Markov chain that samples from $p_{\text{model}}(\mathbf{x})$ is to use **Gibbs sampling**, in which sampling from $T(\mathbf{x}' | \mathbf{x})$ is accomplished by selecting one variable x_i and sampling it from p_{model} conditioned on its neighbors in the undirected graph \mathcal{G} defining the structure of the energy-based model. It is also possible to sample several variables at the same time so long as they are conditionally independent given all of their neighbors. As shown in the RBM example in section 16.7.1, all of the hidden units of an RBM may be sampled simultaneously because they are conditionally independent from each other given all of the visible units. Likewise, all of the visible units may be sampled simultaneously because they are conditionally independent from each other given all of the hidden units. Gibbs sampling approaches that update many variables simultaneously in this way are called **block Gibbs sampling**.

Alternate approaches to designing Markov chains to sample from p_{model} are possible. For example, the Metropolis-Hastings algorithm is widely used in other disciplines. In the context of the deep learning approach to undirected modeling, it is rare to use any approach other than Gibbs sampling. Improved sampling techniques are one possible research frontier.

17.5 The Challenge of Mixing between Separated Modes

The primary difficulty involved with MCMC methods is that they have a tendency to **mix** poorly. Ideally, successive samples from a Markov chain designed to sample