



Figure 19.2: The graph structure of a binary sparse coding model with four hidden units. (Left) The graph structure of $p(\mathbf{h}, \mathbf{v})$. Note that the edges are directed, and that every two hidden units are co-parents of every visible unit. (Right) The graph structure of $p(\mathbf{h} | \mathbf{v})$. In order to account for the active paths between co-parents, the posterior distribution needs an edge between all of the hidden units.

$$= \frac{\sum_{\mathbf{h}} p(\mathbf{v} | \mathbf{h}) \frac{\partial}{\partial b_i} p(\mathbf{h})}{p(\mathbf{v})} \quad (19.25)$$

$$= \sum_{\mathbf{h}} p(\mathbf{h} | \mathbf{v}) \frac{\frac{\partial}{\partial b_i} p(\mathbf{h})}{p(\mathbf{h})} \quad (19.26)$$

$$= \mathbb{E}_{\mathbf{h} \sim p(\mathbf{h} | \mathbf{v})} \frac{\partial}{\partial b_i} \log p(\mathbf{h}). \quad (19.27)$$

This requires computing expectations with respect to $p(\mathbf{h} | \mathbf{v})$. Unfortunately, $p(\mathbf{h} | \mathbf{v})$ is a complicated distribution. See figure 19.2 for the graph structure of $p(\mathbf{h}, \mathbf{v})$ and $p(\mathbf{h} | \mathbf{v})$. The posterior distribution corresponds to the complete graph over the hidden units, so variable elimination algorithms do not help us to compute the required expectations any faster than brute force.

We can resolve this difficulty by using variational inference and variational learning instead.

We can make a mean field approximation:

$$q(\mathbf{h} | \mathbf{v}) = \prod_i q(h_i | \mathbf{v}). \quad (19.28)$$

The latent variables of the binary sparse coding model are binary, so to represent a factorial q we simply need to model m Bernoulli distributions $q(h_i | \mathbf{v})$. A natural way to represent the means of the Bernoulli distributions is with a vector $\hat{\mathbf{h}}$ of probabilities, with $q(h_i = 1 | \mathbf{v}) = \hat{h}_i$. We impose a restriction that \hat{h}_i is never equal to 0 or to 1, in order to avoid errors when computing, for example, $\log \hat{h}_i$.

We will see that the variational inference equations never assign 0 or 1 to \hat{h}_i