

$$= \frac{1}{Z'} \exp \left\{ \sum_{j=1}^{n_h} c_j h_j + \sum_{j=1}^{n_h} \mathbf{v}^\top \mathbf{W}_{:,j} \mathbf{h}_j \right\} \quad (20.10)$$

$$= \frac{1}{Z'} \prod_{j=1}^{n_h} \exp \left\{ c_j h_j + \mathbf{v}^\top \mathbf{W}_{:,j} \mathbf{h}_j \right\} \quad (20.11)$$

Since we are conditioning on the visible units  $\mathbf{v}$ , we can treat these as constant with respect to the distribution  $P(\mathbf{h} \mid \mathbf{v})$ . The factorial nature of the conditional  $P(\mathbf{h} \mid \mathbf{v})$  follows immediately from our ability to write the joint probability over the vector  $\mathbf{h}$  as the product of (unnormalized) distributions over the individual elements,  $h_j$ . It is now a simple matter of normalizing the distributions over the individual binary  $h_j$ .

$$P(h_j = 1 \mid \mathbf{v}) = \frac{\tilde{P}(h_j = 1 \mid \mathbf{v})}{\tilde{P}(h_j = 0 \mid \mathbf{v}) + \tilde{P}(h_j = 1 \mid \mathbf{v})} \quad (20.12)$$

$$= \frac{\exp \{c_j + \mathbf{v}^\top \mathbf{W}_{:,j}\}}{\exp \{0\} + \exp \{c_j + \mathbf{v}^\top \mathbf{W}_{:,j}\}} \quad (20.13)$$

$$= \sigma \left( c_j + \mathbf{v}^\top \mathbf{W}_{:,j} \right). \quad (20.14)$$

We can now express the full conditional over the hidden layer as the factorial distribution:

$$P(\mathbf{h} \mid \mathbf{v}) = \prod_{j=1}^{n_h} \sigma \left( (2\mathbf{h} - 1) \odot (\mathbf{c} + \mathbf{W}^\top \mathbf{v}) \right)_j. \quad (20.15)$$

A similar derivation will show that the other condition of interest to us,  $P(\mathbf{v} \mid \mathbf{h})$ , is also a factorial distribution:

$$P(\mathbf{v} \mid \mathbf{h}) = \prod_{i=1}^{n_v} \sigma \left( (2\mathbf{v} - 1) \odot (\mathbf{b} + \mathbf{W}\mathbf{h}) \right)_i. \quad (20.16)$$

### 20.2.2 Training Restricted Boltzmann Machines

Because the RBM admits efficient evaluation and differentiation of  $\tilde{P}(\mathbf{v})$  and efficient MCMC sampling in the form of block Gibbs sampling, it can readily be trained with any of the techniques described in chapter 18 for training models that have intractable partition functions. This includes CD, SML (PCD), ratio matching and so on. Compared to other undirected models used in deep learning, the RBM is relatively straightforward to train because we can compute  $P(\mathbf{h} \mid \mathbf{v})$