

where  $\mathbf{C}_{\mathbf{x}|\mathbf{h}}^{\text{ss}} = (\mathbf{\Lambda} + \sum_i \mathbf{\Phi}_i h_i - \sum_i \alpha_i^{-1} h_i \mathbf{W}_{:,i} \mathbf{W}_{:,i}^\top)^{-1}$ . The last equality holds only if the covariance matrix  $\mathbf{C}_{\mathbf{x}|\mathbf{h}}^{\text{ss}}$  is positive definite.

Gating by the spike variables means that the true marginal distribution over  $\mathbf{h} \odot \mathbf{s}$  is sparse. This is different from sparse coding, where samples from the model “almost never” (in the measure theoretic sense) contain zeros in the code, and MAP inference is required to impose sparsity.

Comparing the ssRBM to the mcRBM and the mPoT models, the ssRBM parametrizes the conditional covariance of the observation in a significantly different way. The mcRBM and mPoT both model the covariance structure of the observation as  $(\sum_j h_j^{(c)} \mathbf{r}^{(j)} \mathbf{r}^{(j)\top} + \mathbf{I})^{-1}$ , using the activation of the hidden units  $h_j > 0$  to enforce constraints on the conditional covariance in the direction  $\mathbf{r}^{(j)}$ . In contrast, the ssRBM specifies the conditional covariance of the observations using the hidden spike activations  $h_i = 1$  to pinch the precision matrix along the direction specified by the corresponding weight vector. The ssRBM conditional covariance is very similar to that given by a different model: the product of probabilistic principal components analysis (PoPPCA) (Williams and Agakov, 2002). In the overcomplete setting, sparse activations with the ssRBM parametrization permit significant variance (above the nominal variance given by  $\mathbf{\Lambda}^{-1}$ ) only in the selected directions of the sparsely activated  $h_i$ . In the mcRBM or mPoT models, an overcomplete representation would mean that to capture variation in a particular direction in the observation space requires removing potentially all constraints with positive projection in that direction. This would suggest that these models are less well suited to the overcomplete setting.

The primary disadvantage of the spike and slab restricted Boltzmann machine is that some settings of the parameters can correspond to a covariance matrix that is not positive definite. Such a covariance matrix places more unnormalized probability on values that are farther from the mean, causing the integral over all possible outcomes to diverge. Generally this issue can be avoided with simple heuristic tricks. There is not yet any theoretically satisfying solution. Using constrained optimization to explicitly avoid the regions where the probability is undefined is difficult to do without being overly conservative and also preventing the model from accessing high-performing regions of parameter space.

Qualitatively, convolutional variants of the ssRBM produce excellent samples of natural images. Some examples are shown in figure 16.1.

The ssRBM allows for several extensions. Including higher-order interactions and average-pooling of the slab variables (Courville *et al.*, 2014) enables the model to learn excellent features for a classifier when labeled data is scarce. Adding a