

Exercise Sheet 15

Exercise 1: RBM with Ternary Hidden Units (20 + 10 P)

We consider a variant of the restricted Boltzmann machine with ternary hidden units $\mathbf{h} \in \{-1, 0, 1\}^H$. Input features remain binary, i.e. $\mathbf{x} \in \{0, 1\}^d$, like for the classical RBM. The probability model is given by:

$$p(\mathbf{x}, \mathbf{h} | \theta) = \frac{1}{\mathcal{Z}} \exp \left(\sum_{j=1}^H \mathbf{w}_j^\top \mathbf{x} \cdot h_j + \sum_{j=1}^H h_j b_j \right)$$

where $\theta = (\mathbf{w}_j, b_j)_{j=1}^H$ are the parameters of the model, and where \mathcal{Z} is the partition function that normalizes the probability distribution to 1.

- (a) Show that this modified RBM can also be expressed as a product of experts

$$p(\mathbf{x} | \theta) = \frac{1}{\mathcal{Z}} \prod_{j=1}^H g_j(\mathbf{x}, \theta_j),$$

with

$$g_j(\mathbf{x}, \theta_j) = 1 + 2 \cosh(\mathbf{w}_j^\top \mathbf{x} + b_j),$$

where \cosh is the hyperbolic cosine function.

- (b) Show that the gradient of the log-likelihood assigned to some data point \mathbf{x}_n by the modified RBM has the form

$$\begin{aligned} \forall_{j=1}^H : \frac{\partial \log p(\mathbf{x}_n | \theta)}{\partial \mathbf{w}_j} &= \mathbf{x}_n \cdot \sigma(\mathbf{w}_j^\top \mathbf{x}_n + b_j) - \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x} | \theta)} [\mathbf{x} \cdot \sigma(\mathbf{w}_j^\top \mathbf{x} + b_j)] \\ \forall_{j=1}^H : \frac{\partial \log p(\mathbf{x}_n | \theta)}{\partial b_j} &= \sigma(\mathbf{w}_j^\top \mathbf{x}_n + b_j) - \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x} | \theta)} [\sigma(\mathbf{w}_j^\top \mathbf{x} + b_j)] \end{aligned}$$

where $\sigma(t) = \frac{\sinh(t)}{0.5 + \cosh(t)}$.

Exercise 2: Product of Gaussian Mixture Models (20 + 10 P)

Consider the product of experts:

$$p(\mathbf{x} | \theta) = \frac{1}{\mathcal{Z}} \prod_{j=1}^H g_j(\mathbf{x}, \theta_j)$$

where each expert is a Gaussian mixture model in d -dimensions, and where each element of the mixture is Gaussian with identity covariance:

$$\forall_{j=1}^H : g_j(\mathbf{x}, \theta_j) = \sum_{k=1}^C \alpha_{jk} \frac{1}{(2\pi)^{d/2}} \exp \left(-\frac{1}{2} \|\mathbf{x} - \boldsymbol{\mu}_{jk}\|^2 \right).$$

- (a) Show that $p(\mathbf{x} | \theta)$ can be rewritten as a mixture of C^H elements, where each mixture element (indexed by the vector $\mathbf{k} \in \{1, \dots, C\}^H$) has center

$$\mathbf{m}_{\mathbf{k}} = \frac{1}{H} \sum_{j=1}^H \boldsymbol{\mu}_{jk_j}.$$

- (b) Give the centers $\mathbf{m}_{\mathbf{k}}$ of the mixture model equivalent to a product of two mixture models, where each mixture model in the product has 2 elements, where the first mixture has the two-dimensional centers $\boldsymbol{\mu}_{11} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ and $\boldsymbol{\mu}_{12} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$, and where the second mixture has the two-dimensional centers $\boldsymbol{\mu}_{21} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$ and $\boldsymbol{\mu}_{22} = \begin{pmatrix} 0 \\ 4 \end{pmatrix}$.

Exercise 3: Programming (40 P)

Download the programming files on ISIS and follow the instructions.