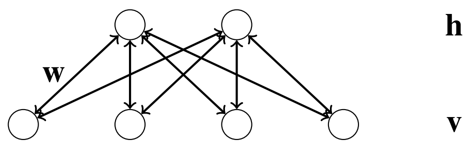
Dynamical Systems and Deep Learning - Exercises 4

1. Consider the following RBM:



with weights:

$$w = \begin{bmatrix} -0.5 & -0.9 & 0.3 & 0.1 \\ 0.6 & -0.1 & 0.2 & -0.3 \end{bmatrix}$$

(so that w_{ij} is the weight between hidden unit i and visible unit j);
hidden biases:

$$c = \begin{bmatrix} 0.1 \\ 0.3 \end{bmatrix}$$

and visible biases:

$$b = \begin{bmatrix} 0 \\ 0.2 \\ 0.1 \\ 0.05 \end{bmatrix}$$

The following training vector is presented to the network:

$$v = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

NB: for the purpose of this exercise, you should activate units deterministically according to:

$$h_i = \begin{cases} 1 & \text{if } \sigma(c_i + \sum_j w_{ij} v_j) > 0.5 \text{ (i.e, if } c_i + \sum_j w_{ij} v_j > 0) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$v_j = \begin{cases} 1 & \text{if } \sigma(b_j + \sum_i w_{ij} h_i) > 0.5 \text{ (i.e, if } b_j + \sum_i w_{ij} h_i > 0) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $\sigma(x)$ is the logistic function:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

You should use 1-step contrastive divergence.

(i) Compute the positive phase for w, b and c.

(ii) Compute the negative phase for w, b and c.

(iii) Compute the new weights, hidden biases and visible biases for learning rate $\alpha = 0.1$.

2. In this exercises, you will show that the joint probability distribution of visible and hidden units in an RBM satisfies the detailed balanced condition with respect to the transitional probability used in block Gibbs sampling.

(i) Show that in an RBM with m visible units V_j, ($j = 1, \ldots, m$), and n hidden units H_i, ($i = 1, \ldots, n$), we have:

$$p(H_i = h_i | v) = \frac{e^{\sum_{j=1}^m w_{ij} v_j h_i + c_i h_i}}{1 + e^{\sum_{j=1}^m w_{ij} v_j + c_i}} \tag{4}$$

where $h_i \in \{0, 1\}$. Similarly, show that

$$p(V_j = v_j | h) = \frac{e^{\sum_{i=1}^n w_{ij} v_j h_i + b_j v_j}}{1 + e^{\sum_{i=1}^n w_{ij} h_i + b_j}} \tag{5}$$

where $v_j \in \{0, 1\}$.

(ii) Use these results to show that the joint probability distribution

$$p(v,h) = \frac{e^{-E(v,h)}}{Z}, \qquad \text{where } Z = \sum_{v\in\{0,1\}^m, h\in\{0,1\}^n} e^{-E(v,h)}$$

satisfies the detailed balanced condition with respect to the transitional probability for block Gibbs sampling.

(iii) Deduce that

$$p(v,h) = \frac{e^{-E(v,h)}}{Z}, \qquad \text{where } Z = \sum_{v\in\{0,1\}^m, h\in\{0,1\}^n} e^{-E(v,h)}$$

is the stationary distribution of the Gibbs sampling.