

## I. Pencil-and-Paper

Suppose  $(V, H)$  denote the visible and hidden random variable which takes values  $(v \in \{0, 1\}^m, h \in \{0, 1\}^n)$ . And the joint probability is  $p(v, h; \theta) = \frac{1}{Z} e^{-E(v, h; \theta)}$ , where  $E$  is the energy function:

$$\begin{aligned} E(v, h; \theta) &= - \sum_{i=1}^n \sum_{j=1}^m w_{ij} h_i v_j - \sum_{j=1}^m b_j v_j - \sum_{i=1}^n c_i h_i \\ &= -(\mathbf{v}^T \mathbf{W} \mathbf{h} + \mathbf{v}^T \mathbf{b} + \mathbf{h}^T \mathbf{c}) \end{aligned} \quad (1)$$

where  $Z = \sum_v \sum_h e^{-E(v, h; \theta)}$  and  $\theta = (W, b, c)$ .

1. Find  $p(v|h, \theta)$  and  $\mathbb{E}(v|h, \theta)$ .

From the structure of RBM, it means that the hidden variables are independent given the state of the visible variables and the visible variables are independent given the state of the hidden variables. In this way,

$$\underline{p(v|h, \theta) = \prod_{i=1}^n p(v_i|h, \theta)} \quad (2)$$

So, we also have:

$$\mathbb{E}(v|h, \theta) = \text{sigmoid}(W^T h + b) \quad (3)$$

For  $p(v_j|h)$ ,

$$p(v_j|h) = \frac{p(v_j, h)}{p(h)} = \frac{p(v_j, h)}{p(v_j = 0, h) + p(v_j = 1, h)} \quad (4)$$

Since that  $p(v_j = 0, h) = \frac{1}{Z} e^{-E(v_j=0, h; \theta)}$  and  $p(v_j = 1, h) = \frac{1}{Z} e^{-E(v_j=1, h; \theta)}$ . From equation (1), we can think  $E(v, h; \theta) = v_j \alpha_j + \beta$ , where  $\beta$  doesn't contain  $v_j$ . In this way, we have:

$$\begin{aligned} p(v_j|h) &= \frac{p(v_j, h)}{p(v_j = 0, h) + p(v_j = 1, h)} \\ &= \frac{\exp(-v_j \alpha_j - \beta)}{\exp(-\beta) + \exp(-\alpha_j - \beta)} \\ &= \frac{\exp(-v_j \alpha_j)}{1 + \exp(-\alpha_j)} \end{aligned} \quad (5)$$

So, we have

$$p(v_j = 1|h) = \frac{\exp(-\alpha_j)}{1 + \exp(-\alpha_j)} = \frac{1}{1 + \exp(\alpha_j)} \quad (6)$$

and

$$p(v_j = 0|h) = \frac{1}{1 + \exp(-\alpha_j)} = 1 - \frac{1}{1 + \exp(\alpha_j)} \quad (7)$$

From equation (1), we can see that  $\alpha_j = -\sum_{i=1}^n w_{ij}h_i - b_j$ , so we can see that:

$$\underline{p(v_j = 1|h) = \text{sigmoid}(-\alpha_j) = \text{sigmoid}(\sum_{i=1}^n w_{ij}h_i + b_j)} \quad (8)$$

and

$$\underline{p(v_j = 0|h) = 1 - \text{sigmoid}(-\alpha_j) = 1 - \text{sigmoid}(\sum_{i=1}^n w_{ij}h_i + b_j)} \quad (9)$$

From equation (8), we can see that:

$$\mathbb{E}(v_j|h, \theta) = 1 \cdot p(v_j = 1|h, \theta) + 0 \cdot p(v_j = 0|h, \theta) = \text{sigmoid}\left(\sum_{i=1}^n w_{ij}h_i + b_j\right) \quad (10)$$

In the vector format:

$$\underline{\mathbb{E}(v|h, \theta) = \text{sigmoid}(W^T h + b)} \quad (11)$$

2. Find  $p(h|v, \theta)$  and  $\mathbb{E}(h|v, \theta)$ .
3. Compute  $\frac{\partial \mathcal{L}(D|\theta)}{\partial W_{ij}}$ ,  $\frac{\partial \mathcal{L}(D|\theta)}{\partial b_j}$  and  $\frac{\partial \mathcal{L}(D|\theta)}{\partial c_i}$
4. Contrastive divergence

## II. Code-from-Scratch

### 1. Methods

### 2. Results

### **III. TensorFlow**

#### **1. Methods**

#### **2. Results**