Posons

$$Z = \sum_x \exp(-H(x, \theta))$$

1. Montrons que:

$$\frac{\partial}{\partial w_{i,a}} \log \mathcal{L} = \mathbb{E}_{x\sim \mathrm{D}}\left[v_i h_a\right] - \mathbb{E}_{x\sim \mathrm{RBM}}\left[v_i h_a\right]$$

On part de l'expression de la log-vraisemblance:

$$\log(\mathcal{L}) = \frac{1}{N_S}\sum_k \log p(\vec{v}^{(k)})$$

Le gradient (par rapport à $w_{i,a}$) de la log-vraisemblance est donc:

$$\begin{aligned}
\frac{\partial}{\partial w_{i,a}} \log(\mathcal{L}) &= \frac{\partial}{\partial w_{i,a}}\left(\frac{1}{N_S}\sum_k \log p(\vec{v}^{(k)})\right)\\
&= \frac{1}{N_S}\sum_k \frac{\partial}{\partial w_{i,a}} \log p(\vec{v}^{(k)})\\
&= \frac{1}{N_S}\sum_k \frac{\partial}{\partial w_{i,a}} \left(-H(\vec{v}^{(k)}, \vec{h}) - \log Z\right)
\end{aligned}$$

On regarde la partie positive :

$$\begin{aligned}
&= \frac{1}{N_s}\sum_{k=1}^{N_s}\left(-\frac{\partial}{\partial w_{i,a}} H(\vec{v}^{(k)}, \vec{h}^{(k)})\right)\\
&= \frac{1}{N_s}\sum_{k=1}^{N_s}\left(\frac{\partial}{\partial w_{i,a}}\left(\sum_{i=1}^{N_v}\sum_{a=1}^{N_h} v_i^{(k)} w_{i,a} h_a^{(k)}\right)\right)\\
&= \frac{1}{N_s}\sum_{k=1}^{N_s}\sum_{i=1}^{N_v}\sum_{a=1}^{N_h} v_i^{(k)} h_a^{(k)}\\
&= \sum_{k=1}^{N_s} v_i^{(k)} h_a^{(k)} \cdot \frac{1}{N_s}\\
&= \mathbb{E}_{x\sim D}\left[v_i h_a\right]
\end{aligned}$$

On regarde la partie négative:

$$\begin{aligned}
\frac{1}{N_s}\sum_k \frac{\partial}{\partial w_{i,a}}(-\log(Z)) &= \frac{1}{N_s}\sum_k \frac{\partial}{\partial w_{i,a}}\left(-\log\left(\sum_k \exp\left(-H(\vec{v}^{(k)}, \vec{h})\right)\right)\right)\\
&= \frac{1}{N_s}\sum_k\left(\frac{-\frac{\partial}{\partial w_{i,a}}\exp\left(-H(\vec{v}^{(k)}, \vec{h})\right)}{\sum_k \exp\left(-H(\vec{v}^{(k)}, \vec{h})\right)}\right)\\
&= \frac{1}{N_s}\sum_k\left(\frac{\left(\sum_k \sum_{i=1}^{N_v}\sum_{a=1}^{N_h} v_i^{(k)} h_a\right)\cdot \exp\left(-H(\vec{v}^{(k)}, \vec{h})\right)}{\sum_k \exp\left(-H(\vec{v}^{(k)}, \vec{h})\right)}\right)\\
&= \frac{1}{N_s}\sum_k v_i^{(k)} h_a \cdot N_s \cdot \left(\frac{\exp\left(-H(\vec{v}^{(k)}, \vec{h})\right)}{\sum_k \exp\left(-H(\vec{v}^{(k)}, \vec{h})\right)}\right)\\
&= \sum_k v_i^{(k)} h_a \cdot P_{x\sim \mathrm{RBM}}\left(\vec{v}^{(k)}\right)\\
&= \mathbb{E}_{x\sim \mathrm{RBM}}\left[v_i h_a\right]
\end{aligned}$$

On conclut finalement

$$\frac{\partial}{\partial \omega_{i,a}} \log \mathcal{L} = \mathbb{E}_{X \sim P_D}\left[v_i h_a\right] - \mathbb{E}_{X \sim P_{\mathrm{RBM}}}\left[v_i h_a\right]$$

2. En outre, on a :

$$p(v,h) = \frac{1}{Z} \exp\left(\sum_i \theta_i v_i + \sum_a \eta_a h_a + \sum_{i,a} w_{ia} v_i h_a\right)$$

$$Z = \sum_{\boldsymbol{v},\boldsymbol{h}} \exp\left(\sum_i \theta_i v_i + \sum_a \eta_a h_a + \sum_{i,a} w_{ia} v_i h_a\right)$$

donc il suit

$$\mathbb{P}(h \mid v) = \frac{p(v,h)}{\sum_h p(v,h)} = \frac{1}{Z} \exp\left(\sum_i \theta_i v_i + \sum_a \eta_a h_a + \sum_{i,a} w_{ia} v_i h_a\right) \times \frac{1}{\sum_h p(v,h)}$$

Montrons que

$$\mathbb{P}(h \mid v) = \prod_a \mathbb{P}(h_a \mid v)$$

En particulier, la somme du dénominateur donne :

$$\begin{aligned}
\sum_h p(v,h) &= \frac{1}{Z} \sum_h \exp\left[-\mathcal{H}(v,h)\right] \\
&= \frac{1}{Z} \sum_h \exp\left[\sum_a h_a \eta_a + \sum_i v_i \theta_i + \sum_{ia} w_{ia} v_i h_a\right] \\
&= \frac{1}{Z} \exp\left[\sum_i v_i \theta_i\right] \sum_h \exp\left[\sum_a h_a \left(\eta_a + \sum_i w_{ia} v_i\right)\right] \\
&= \frac{1}{Z} \exp\left[\sum_i v_i \theta_i\right] \prod_a \sum_{h_a} \exp\left[h_a \left(\eta_a + \sum_i w_{ia} v_i\right)\right] \\
&= \frac{1}{Z} \exp\left[\sum_i v_i \theta_i\right] \prod_a \sum_{h_a \in \{0,1\}} \exp\left[h_a \left(\eta_a + \sum_i w_{ia} v_i\right)\right] \\
&= \frac{1}{Z} \exp\left[\sum_i v_i \theta_i\right] \prod_a \left(1 + \exp\left[\eta_a + \sum_i w_{ia} v_i\right]\right)
\end{aligned}$$

en injectant dans la formule :

$$
\begin{aligned}
\mathbb{P}(h \mid v) &= \frac{\frac{1}{Z}\exp\big[-H(v,h)\big]}{\sum_v p(v,h)} \\
&= \frac{1}{Z} \times Z \times \frac{\exp\Big[\sum_a h_a\eta_a + \sum_i v_i\theta_i + \sum_{i,a} w_{ia}v_ih_a\Big]}{\exp\left[\sum_i v_i\theta_i\right]\prod_a\left(1+\exp\left[\eta_a + \sum_i w_{ia}v_i\right]\right)} \\
&= \frac{\exp\left[\sum_i v_i\theta_i\right]\exp\Big[\sum_a h_a\eta_a + \sum_{i,a} w_{ia}v_ih_a\Big]}{\exp\left[\sum_i v_i\theta_i\right]\prod_a\left(1+\exp\left[\eta_a + \sum_i w_{ia}v_i\right]\right)} \\
&= \frac{\exp\Big[\sum_a h_a\eta_a + \sum_{i,a} w_{ia}v_ih_a\Big]}{\prod_a\left(1+\exp\left[\eta_a + \sum_i w_{ia}v_i\right]\right)} \\
&= \prod_a \frac{\exp\left[h_a\left(\eta_a + \sum_i w_{ia}v_i\right)\right]}{1+\exp\left[\eta_a + \sum_i w_{ia}v_i\right]} \\
&= \prod_a \mathbb{P}(h_a \mid v)
\end{aligned}
$$

Notamment,

$$
\begin{aligned}
\mathbb{P}(h_a = 1 \mid v) &= \frac{\exp(\eta_a + \sum_i w_{ia}v_i)}{1+\exp(\eta_a + \sum_i w_{ia}v_i)} \\
&= \sigma\left(\eta_a + \sum_i w_{ia}v_i\right)
\end{aligned}
$$

où σ désigne la fonction sigmoïde.
De même,

$$
\begin{aligned}
\mathbb{P}(v_i = 1 \mid h) &= \frac{\exp(\theta_i + \sum_a w_{ia}h_a)}{1+\exp(\theta_i + \sum_a w_{ia}h_a)} \\
&= \sigma\left(\theta_i + \sum_a w_{ia}h_a\right)
\end{aligned}
$$

3. On cherche à initialiser les biais des unités visibles θ_i de façon à ce que la probabilité marginale $P(v_i = 1)$ soit égale à la moyenne empirique observée dans le jeu de données, notée $\mathbb{E}_{x\sim\mathcal{D}}[v_i]$.

Dans une RBM, on a :

$$P(v_i = 1) = \sigma(\theta_i) = \frac{1}{1+e^{-\theta_i}}$$

et on souhaite que :

$$\sigma(\theta_i) = \mathbb{E}_{x\sim\mathcal{D}}[v_i]$$

On en déduit alors :

$$\theta_i = \sigma^{-1}(\mathbb{E}_{x\sim\mathcal{D}}[v_i]) = \log\left(\frac{\mathbb{E}_{x\sim\mathcal{D}}[v_i]}{1-\mathbb{E}_{x\sim\mathcal{D}}[v_i]}\right)$$

Ce qui justifie l'initialisation :

$$\theta_i = \log\left(\mathbb{E}_{x\sim\mathcal{D}}[v_i]\right) - \log\left(1-\mathbb{E}_{x\sim\mathcal{D}}[v_i]\right)$$