RBM Training

Here we derive the rule for training an RBM.

Let $q(v)$ be the probability for v to be sampled from the data.

Let $p(v) \equiv \frac{e^{-F(v)}}{Z} = \frac{\sum_h e^{-E(v,h)}}{Z}$ $\quad Z = \sum_v e^{-F(v)}$

Then $F(v) = -\ln\left(\sum_h e^{-E(v,h)}\right)$

$$\frac{\partial F(v)}{\partial W} = + \frac{\sum_h e^{-E(v,h)}}{\sum_h e^{-E(v,h)}} \frac{\partial E(v,h)}{\partial W}$$

$$= \sum_h p(h|v) \frac{\partial E(v,h)}{\partial W}$$

We want to maximize the KL-divergence:

$$KL = \sum_v q(v) \ln(q(v)) - q(v) \ln(p(v))$$

This term doesn't depend on the weights or biases

$$\left[q(v) \ln p(v)\right] = q \ln\left(\frac{e^{-F}}{Z}\right) = -qF - q \ln Z$$

$$= -qF - q \ln\left(\sum_v e^{-F(v)}\right)$$

$$\sum_v \frac{\partial}{\partial W}\left(-qF - q\ln\left(\sum_v e^{-F(v)}\right)\right)$$

$$= \sum_v -q \frac{\partial F}{\partial W} + q \frac{\sum_v e^{-F(v)}}{\sum_{v'} e^{-F(v')}} \frac{\partial F(v)}{\partial W}$$

$p(v)$

$$= \sum_v -q \frac{\partial F(v)}{\partial w} + \sum_v q \sum_{v'} p(v') \frac{\partial F(v')}{\partial w}$$

$$\sum_v -q(v) \frac{\partial F(v)}{\partial w} + \sum_{v'} p(v') \frac{\partial F(v')}{\partial w}$$

$$\underbrace{\sum_v -q(v) \sum_h p(h|v) \frac{\partial E(v,h)}{\partial w}}_{} + \sum_{v'} p(v') \sum_h p(h|v') \frac{\partial E(v',h)}{\partial w}$$

$$\left\langle -\frac{\partial E(v,h)}{\partial w} \right\rangle_{q(v),\, p(h|v)}$$

→ This is the stochastic thing you are doing in CD_k

$$\left\langle -\sum_i h_i (Wv)_i \right\rangle_{q(v)\, p(h|v)} = -\left\langle h_1 (Wv)_1 \right\rangle_{q(v)\, p(h|v)} + -\left\langle h_2 (Wv)_2 \right\rangle_{q(v)\, p(h|v)} + \dots$$

$$= \left\langle p(h_1 = 1|v) (Wv)_1 \right\rangle_{q(v)} + \left\langle p(h_2=1) (Wv)_2 \right\rangle_{q(v)} + \dots$$

$$= \sum_i \left\langle \underline{p(h_i = 1|v) (Wv)_i} \right\rangle_{q(v)}$$

→ a more accurate analytic choice