7) Monte Carlo Chains? → Gibbs Sampling

~~ϕ~~ $\hat{x} = \frac{1}{n} \sum_{i}^{n} p(x) f(x)$

Transformation probabilities $T(x'|x)$.

Patricks

Deep Learning Part 2.

18/06/2021

$$P(v|h) = \prod_{i=1}^{n_v} \sigma\big((2v-1) \odot (b + Wh)\big)_i$$

In RBMs:

$$P(v,h) = \frac{1}{Z} \exp(-E(v,h))$$

$$E(v,h) = -b^T v - c^T h - v^T W h$$

$$P(v|h) = \frac{P(v,h)}{P(h)}$$

$$= \frac{1}{P(h)} \frac{1}{Z} \exp(+b^T v + c^T h + v^T W h)$$

→ We observe h so terms dependent on h are constants.

$$= \frac{1}{Z'} \exp(b^T v + v^T W h)$$

$$= \frac{1}{Z'} \exp\left(b^T v + v^T W h \right)$$

$$= \frac{1}{Z'} \exp\left(\sum_{i=1}^{n_v} b_i v_i + v_i W h \right)$$

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

$$= \frac{1}{Z'} \prod_{i=1}^{n_v} \exp\left(b_i v_i + v_i W h \right)$$

$$= \frac{1}{Z'} \prod_{i=1}^{n_v} \exp(v_i) \exp(b_i + W h)$$

- Normalise:

$$P(v_i = 1 | h) = \frac{\tilde{P}(v_i = 1 | h)}{\tilde{P}(v_i = 0 | h) + \tilde{P}(v_i = 1 | h)}$$

$$= \frac{\exp(b_i + W h)}{\exp(0) + \exp(b_i + W h)}$$

$$= \sigma(b_i + W h)$$

Thus! for visible layers

$$\boxed{P(v | h) = \prod_{i=1}^{n_v} \sigma\left((2v - 1) \odot (b + W h) \right)_i}$$

Patrick — Deep Learning Part 2 — 18/06/21

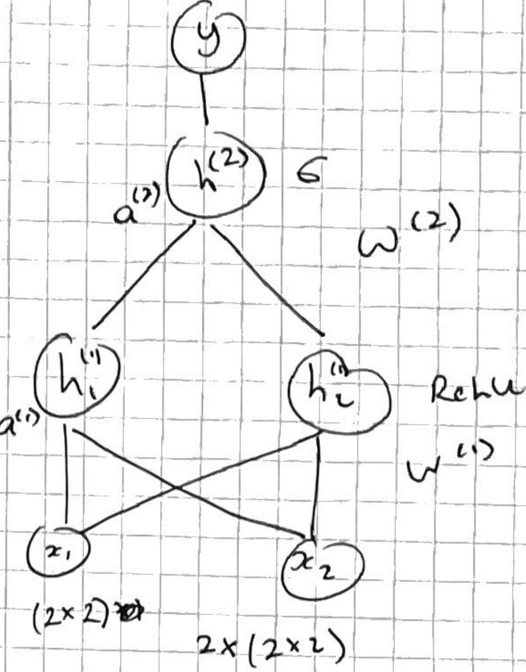
2) $h^{(1)} = W^{(1)}x + b^{(1)}$

$a^{(1)} = \mathrm{ReLU}(h^{(1)})$

$h^{(2)} = (w^{(2)})^T a^{(1)} + b^{(2)}$

$y = \sigma(h^{(2)})$ $\quad y = a^{(2)}$

$J = \frac{1}{2}(t-y)^2$



$1\times 1 \quad 1\times 2 \quad \cancel{1\times(2\times 1)}$ ✓ ✓

$$\frac{\partial J}{\partial W^{(1)}} = \frac{\partial J}{\partial y}\frac{\partial y}{\partial a^{(2)}}\frac{\partial a^{(2)}}{\partial h^{(2)}}\frac{\partial h^{(2)}}{\partial a^{(1)}}\frac{\partial a^{(1)}}{\partial h^{(1)}}\frac{\partial h^{(1)}}{\partial W^{(1)}}$$

1×1 (1) (2) (3) (4) (5) (6)

(1) $\frac{\partial J}{\partial y} = -(t-y)$

(3) $\frac{\partial a^{(2)}}{\partial h^{(2)}} = (h^{(2)})(1-h^{(2)})$

$\sigma(h^{(2)})\cdot(1-\sigma(h^{(2)}))$

(2) $\frac{\partial y}{\partial a^{(2)}} = 1$

(4) $\frac{\partial h^{(2)}}{\partial a^{(1)}} = w^{(2)}$

(5) $\frac{\partial a^{(1)}}{\partial h^{(1)}} = \begin{cases} 1, & h_i \geq 0 \\ 0, & h_i < 0 \end{cases}$ $\quad \rightarrow h^{(1)} = \begin{bmatrix} 0.75 & 0.5 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -0.5 \end{bmatrix} + 1$

$= \begin{bmatrix} 0.75 + 0.25 \\ -0.5 \end{bmatrix} = \begin{bmatrix} 1 \\ -0.5 \end{bmatrix}$

(6) $\frac{\partial h^{(1)}}{\partial W^{(1)}} = x$ $\quad h^{(2)} = \begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} - 1 = 0$

$$\frac{\partial J}{\partial W^{(1)}} = -(t-y)\,\sigma(h^{(2)})\left(1-\sigma(h^{(2)})\right) w^{(2)} \frac{\partial a^{(1)}}{\partial h^{(1)}}^T x$$

$$= -(t-y)\,\frac{1}{4}\, w^{(2)} \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0.5 \end{bmatrix} =$$

$$= -(t-y)\,\frac{1}{4}\,w^{(2)}\,\frac{\partial a^{(1)}}{\partial h^{(1)}}\,\frac{\partial h^{(1)}}{\partial w^{(1)}}$$

$$= \underbrace{-\frac{1}{8}\begin{bmatrix} 1 & 1 \end{bmatrix}}_{1\times 2}\;\underbrace{\frac{\partial a^{(1)}}{\partial h^{(1)}}}_{2\times 2}\;\underbrace{\frac{\partial h^{(1)}}{\partial w^{(1)}}}_{2\times(2\times 2)} = \begin{bmatrix} 2\times 2 \end{bmatrix}$$

$$= \frac{\partial h^{(1)}}{\partial W^{(1)}} = \begin{bmatrix} \overbrace{\dfrac{\partial h_1^{(1)}}{\partial w_1^{(1)}}}^{2\times 2} \\[2ex] \dfrac{\partial h_2^{(1)}}{\partial w_2^{(1)}} \end{bmatrix}$$

~~$h = W^{(1)} x + b^{(1)}$~~

$$\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$\begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \end{bmatrix} = \begin{bmatrix} w_{11}x_1 + w_{12}x_2 + b_1 \\ w_{21}x_1 + w_{22}x_2 + b_2 \end{bmatrix}$$

$$\frac{\partial h_1^{(1)}}{\partial W^{(1)}} = \begin{bmatrix} x_1 & x_2 \\ 0 & 0 \end{bmatrix}$$

$$\frac{\partial h_2^{(1)}}{\partial W^{(1)}} = \begin{bmatrix} 0 & 0 \\ x_1 & x_2 \end{bmatrix}$$

$$\frac{\partial a^{(1)}}{\partial b^{(1)}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} -\frac{1}{8} & 0 \end{bmatrix} \begin{bmatrix} 2\times 2 \\ 2\times 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{8}x_1 & -\frac{1}{8}x_2 \\ 0 & 0 \end{bmatrix}$$