Exercise Sheet 6

Machine Learning 2, SS16

June 2, 2016

Mario Tambos, 380599; Viktor Jeney, 348969; Sascha Huk, 321249; Jan Tinapp, 0380549

Exercise 1 - RBM with Ternary Hidden Units

(a) We have to show:

$$\sum_{h\in\{-1,0,1\}^N} e^{x^\top a + x^\top Wh + h^\top b} \quad \propto \quad e^{x^\top a + \sum_{j=1}^N \log(\frac{1}{2} + \cosh(w_j^\top x + b_j)}$$

Proof. Devide both sides by $\exp(x^\top a)$ and write down the multiplication by h more explicitly:

$$\sum_{h\in\{-1,0,1\}^N} e^{\sum_{j=1}^N h_j w_j^\top x + h_j b_j} \quad \propto \quad e^{\sum_{j=1}^N \log(\frac{1}{2} + \cosh(w_j^\top x + b_j)}$$

It follows by laws of exponentiation:

$$\sum_{h\in\{-1,0,1\}^N} \prod_{j=1}^N e^{h_j w_j^\top x + h_j b_j} \quad \propto \quad \cancel{\exp(\log} \prod_{j=1}^N (\frac{1}{2} + \cosh(w_j^\top x + b_j))$$

Because the expression $\exp(h_j w_j^\top x + h_j b_j)$ only depends on the j'th component of h, we can rewrite the sum and product to get:

$$\prod_{j=1}^N \sum_{h_j\in\{-1,0,1\}} e^{h_j w_j^\top x + h_j b_j} \quad \propto \quad \prod_{j=1}^N \frac{(1 + 2\cosh(w_j^\top x + b_j))}{2}$$

Unfold the three cases on the left side and pull out the constant factor on the right side:

$$\prod_{j=1}^N \left(e^{w_j^\top x + b_j} + e^{-w_j^\top x - b_j} + 1\right) \quad \propto \quad \frac{1}{2^N} \prod_{j=1}^N (1 + 2\cosh(w_j^\top x + b_j))$$

Use definition of cosh on the left side:

$$\prod_{j=1}^N \left(2\cosh(w_j^\top x + b_j) + 1\right) \quad \propto \quad \frac{1}{2^N} \prod_{j=1}^N (1 + 2\cosh(w_j^\top x + b_j))$$

□

(c)

The independence model is $(x_i \perp\!\!\!\perp x_j \mid h)$ and $(h_i \perp\!\!\!\perp h_j \mid x)$ for all $i \neq j$:

$$p(h \mid x) \overset{\text{Model}}{=} \prod_{j=1}^{N} \Pr_j(h_j \mid x)$$

$$p(h \mid x) = \frac{p(h,x)}{p(x)} = \frac{\cancel{\frac{1}{Z}e^{x^\top a}} \prod_j e^{h_j w_j^\top x + h_j b_j}}{\cancel{\frac{1}{Z}e^{x^\top a}} \prod_j (e^{w_j^\top x + b_j} + e^{-w_j^\top x - b_j} + 1)} \overset{\text{Model}}{=} \frac{\prod_j \Pr_j(h_j, x)}{\prod_j \Pr_j(x)} = \prod_j \Pr_j(h_j \mid x)$$

$$\Longrightarrow \Pr_j(h_j \mid x) = \frac{\Pr_j(h_j, x)}{\Pr_j(x)} = \frac{e^{h_j w_j^\top x + h_j b_j}}{\sum_{h_j} e^{h_j w_j^\top x + h_j b_j}} = \frac{e^{h_j w_j^\top x + h_j b_j}}{e^{w_j^\top x + b_j} + e^{-w_j^\top x - b_j} + 1}$$

This is kind of a softmax function.

$$\Pr_j(h_j = 1 \mid x) = \frac{e^{w_j^\top x + b_j}}{e^{w_j^\top x + b_j} + e^{-w_j^\top x - b_j} + 1} = \frac{1}{1 + \exp(-2w_j x - 2b_j) + \exp(-w_j x - b_j)}$$

$$\Pr_j(h_j = -1 \mid x) = \frac{e^{-w_j^\top x - b_j}}{e^{w_j^\top x + b_j} + e^{-w_j^\top x - b_j} + 1} = \frac{1}{\exp(2w_j x + 2b_j) + 1 + \exp(w_j x + b_j)}$$

$$\Pr_j(h_j = 0 \mid x) = \frac{1}{e^{w_j^\top x + b_j} + e^{-w_j^\top x - b_j} + 1} = \frac{1}{\exp(w_j x + b_j) + \exp(-w_j x - b_j) + 1}$$

$$p(x \mid h) \overset{\text{Model}}{=} \prod_{k=1}^{d} \Pr_k(x_k \mid h)$$

$$p(x \mid h) = \frac{p(x,h)}{p(h)} = \frac{\cancel{\frac{1}{Z}e^{h^\top b}} \prod_k e^{x_k a_k + x_k w_k^\top h}}{\cancel{\frac{1}{Z}e^{h^\top b}} \prod_k (1 + e^{a_k + w_k^\top h})} \overset{\text{Model}}{=} \frac{\prod_k \Pr_k(x_k, h)}{\prod_k \Pr_k(h)} = \prod_k \Pr_k(x_k \mid h)$$

$$\Pr(x_k = 1 \mid h) = \frac{e^{a_k + w_k^\top h}}{1 + e^{a_k + w_k^\top h}} = \frac{1}{e^{-a_k - w_k^\top h} + 1} = \text{sigm}(a_k + w_k^\top h)$$

$$\Pr(x_k = 0 \mid h) = \frac{1}{1 + e^{a_k + w_k^\top h}} = \text{sigm}(-a_k - w_k^\top h)$$