

# Exercise Sheet 6

Machine Learning 2, SS16

June 6, 2016

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

## Exercise 1

(a)

Define

$$y := x^T W - b^T$$

$$\begin{aligned} p_\theta(x) &= \sum_{h \in \{-1,0,1\}^N} p(x, h) \\ &= \sum_{h \in \{-1,0,1\}^N} \frac{1}{Z} \exp(yh + x^T a) \\ &= \frac{1}{Z} \exp(x^T a) \sum_{h \in \{-1,0,1\}^N} \exp(yh) \\ &= \frac{1}{Z} \exp(x^T a) \sum_{h \in \{-1,0,1\}^N} \exp\left(\sum_{i=1}^N y_i h_i\right) \\ &= \frac{1}{Z} \exp(x^T a) \sum_{h \in \{-1,0,1\}^N} \prod_{i=1}^N \exp(y_i h_i) \end{aligned}$$

Because the expression  $\exp(y_i h_i)$  only depends on the  $i$ 'th component of  $h$ , we can rewrite the sum and product to get:

$$\begin{aligned} p_\theta(x) &= \frac{1}{Z} \exp(x^T a) \prod_{i=1}^N \sum_{h \in \{-1,0,1\}} \exp(y_i h_i) \\ &= \frac{1}{Z} \exp(x^T a) \exp\left(\log\left(\prod_{i=1}^N \sum_{h \in \{-1,0,1\}} \exp(y_i h_i)\right)\right) \\ &= \frac{1}{Z} \exp(x^T a) \exp\left(\sum_{i=1}^N \log\left(\sum_{h \in \{-1,0,1\}} \exp(y_i h_i)\right)\right) \\ &= \frac{1}{Z} \exp(x^T a) \exp\left(\sum_{i=1}^N \log(1 + e^{y_i} + e^{-y_i})\right) \\ &= \frac{1}{Z} \exp(x^T a) \exp\left(\sum_{i=1}^N \log(1 + 2\cosh(y_i))\right) \\ &= \frac{1}{Z} \exp(x^T a + \sum_{i=1}^N \log(1 + 2\cosh(w_i x - b_i))) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{Z} \exp(x^T a + \sum_{i=1}^N \log(2(\frac{1}{2} + \cosh(w_i x - b_i)))) \\
&= \frac{1}{Z} \exp(x^T a + \sum_{i=1}^N \log(2) + \log(\frac{1}{2} + \cosh(w_i x - b_i))) \\
&= \frac{1}{Z} \exp(N \log(2) + x^T a + \sum_{i=1}^N \log(\frac{1}{2} + \cosh(w_i x - b_i))) \\
&= \frac{1}{Z} 2^N \exp(x^T a + \sum_{i=1}^N \log(\frac{1}{2} + \cosh(w_i x - b_i)))
\end{aligned}$$

With

$$Z' := \frac{2^N}{Z}$$

the desired result follows.

**(b)**

First compute gradients of  $F$ :

$$\begin{aligned}
\nabla_{a_i} F(x) &= \nabla_{a_i} - a^T x = -x_i \\
\nabla_{b_j} F(x) &= - \sum_{k=1}^N \nabla_{b_j} \log(\frac{1}{2} + \cosh(w_k x - b_k)) \\
&= - \nabla_{b_j} \log(\frac{1}{2} + \cosh(w_j x - b_j)) \\
&= - \frac{1}{\frac{1}{2} + \cosh(w_k x - b_k)} \nabla_{b_j} \cosh(w_j x - b_j) \\
&= - \frac{1}{\frac{1}{2} + \cosh(w_k x - b_k)} \sinh(w_j x - b_j) \nabla_{b_j} b_j \\
&= - \frac{\sinh(w_j x - b_j)}{\frac{1}{2} + \cosh(w_k x - b_k)} \\
\nabla_{w_{ij}} F(x) &= - \sum_{k=1}^N \nabla_{w_{ij}} \log(\frac{1}{2} + \cosh(w_k x - b_k)) \\
&= - \nabla_{w_{ij}} \log(\frac{1}{2} + \cosh(w_j x - b_j)) \\
&= - \frac{\sinh(w_j x - b_j)}{\frac{1}{2} + \cosh(w_k x - b_k)} \nabla_{w_{ij}} (w_j^T x) \\
&= - \frac{\sinh(w_j x - b_j)}{\frac{1}{2} + \cosh(w_k x - b_k)} x_i
\end{aligned}$$

Now we can plug the gradients into the expectations:

$$\begin{aligned}
\nabla_{a_i} KL(\hat{p}||p_\theta) &= \langle -x_i \rangle_{\hat{p}} - \langle -x_i \rangle_{p_\theta} \\
&= \\
\nabla_{b_j} KL(\hat{p}||p_\theta) &= \langle -\frac{\sinh(w_j x - b_j)}{\frac{1}{2} + \cosh(w_k x - b_k)} \rangle_{\hat{p}} - \langle -\frac{\sinh(w_j x - b_j)}{\frac{1}{2} + \cosh(w_k x - b_k)} \rangle_{p_\theta} \\
&= \\
\nabla_{w_{ij}} KL(\hat{p}||p_\theta) &= \langle -\frac{\sinh(w_j x - b_j)}{\frac{1}{2} + \cosh(w_k x - b_k)} x_i \rangle_{\hat{p}} - \langle -\frac{\sinh(w_j x - b_j)}{\frac{1}{2} + \cosh(w_k x - b_k)} x_i \rangle_{p_\theta} \\
&=
\end{aligned}$$