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

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

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

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:

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

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

With

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

the desired result follows.

(b)

First compute gradients of F:

$$\bigtriangledown_{a_i} F(x) = \bigtriangledown_{a_i} - a^T x = -x_i$$

$$
\begin{aligned}
\bigtriangledown_{b_j} F(x) &= -\sum_{k=1}^{N} \bigtriangledown_{b_j} \log(\frac{1}{2} + cosh(w_k x - b_k)) \\
&= -\bigtriangledown_{b_j} \log(\frac{1}{2} + cosh(w_j x - b_j)) \\
&= -\frac{1}{\frac{1}{2} + cosh(w_k x - b_k)} \bigtriangledown_{b_j} cosh(w_j x - b_j) \\
&= -\frac{1}{\frac{1}{2} + cosh(w_k x - b_k)} sinh(w_j x - b_j) \bigtriangledown_{b_j} b_j \\
&= -\frac{sinh(w_j x - b_j)}{\frac{1}{2} + cosh(w_k x - b_k)}
\end{aligned}
$$

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

Now we can plug the gradients into the expectations:
To compute the expectations with respect to the empirical distribution, we consider observed data points $x^{(1)}, ..., x^{(n)}$

$$
\begin{aligned}
\bigtriangledown_{a_i} KL(\hat{p}||p_\theta) &= < -x_i >_{\hat{p}} - < -x_i >_{p_\theta} \\
&= \sum_{x\in\{0,1\}^d} x_i p_\theta(x) - < x_i >_{\hat{p}} \\
&= \sum_{x\in\{0,1\}^d} x_i p_\theta(x) - \frac{1}{n}\sum_{k=1}^{n} x_i^{(k)} \\
&= \frac{1}{Z'} \sum_{x\in\{0,1\}^d} x_i \exp(-F_\theta(x)) - \frac{1}{n}\sum_{k=1}^{n} x_i^{(k)}
\end{aligned}
$$

$$
\begin{aligned}
\bigtriangledown_{b_j} KL(\hat{p}||p_\theta) &= < -\frac{sinh(w_j x - b_j)}{\frac{1}{2} + cosh(w_k x - b_k)} >_{\hat{p}} - < -\frac{sinh(w_j x - b_j)}{\frac{1}{2} + cosh(w_k x - b_k)} >_{p_\theta} \\
&= \frac{1}{Z'} \sum_{x \in \{0,1\}^d} \frac{sinh(w_j x - b_j)}{\frac{1}{2} + cosh(w_k x - b_k)} \exp(-F_\theta(x)) - \frac{1}{n} \sum_{k=1}^{n} \frac{sinh(w_j x^{(k)} - b_j)}{\frac{1}{2} + cosh(w_k x^{(k)} - b_k)}
\end{aligned}
$$

$$
\begin{aligned}
\bigtriangledown_{w_{ij}} KL(\hat{p}||p_\theta) &= < -\frac{sinh(w_j x - b_j)}{\frac{1}{2} + cosh(w_k x - b_k)} x_i >_{\hat{p}} - < -\frac{sinh(w_j x - b_j)}{\frac{1}{2} + cosh(w_k x - b_k)} x_i >_{p_\theta} \\
&= \frac{1}{Z'} \sum_{x \in \{0,1\}^d} \frac{sinh(w_j x - b_j)}{\frac{1}{2} + cosh(w_k x - b_k)} x_i \exp(-F_\theta(x)) - \frac{1}{n} \sum_{k=1}^{n} \frac{sinh(w_j x^{(k)} - b_j)}{\frac{1}{2} + cosh(w_k x^{(k)} - b_k)} x_i^{(k)}
\end{aligned}
$$