ECE 544 Fall 2016 Assignment 4

Huan Yan Date: 10/24/2016

Pencil and Paper

1.

$$p(v|h,\theta) = \frac{p(v,h|\theta)}{p(h|\theta)} \tag{1}$$

$$p(h|\theta) = \sum_v p(v,h|\theta) \tag{2}$$

$$p(v|h,\theta) = \frac{p(v,h|\theta)}{\sum_v p(v,h|\theta)} \tag{3}$$

Since each node is independent, the $p(v,h|\theta)$ can be written as:

$$p(v,h|\theta) = \frac{\prod_i^n \prod_j^m e^{-W_{ij}h_i v_j - b_j v_j - c_i h_i}}{\sum_v \sum_h \prod_i^n \prod_j^m e^{-W_{ij}h_i v_j - b_j v_j - c_i h_i}} \tag{4}$$

$$= \frac{\prod_i^n \prod_j^m p(v_j, h_i|\theta)}{\sum_v \sum_h \prod_i^n \prod_j^m p(v_j, h_i|\theta)} \tag{5}$$

$$\tag{6}$$

Then the conditional probability of v can be expressed as:

$$p(v|h,\theta) = \frac{\prod_i^n \prod_j^m p(v_j, h_i|\theta)}{\sum_v \prod_i^n \prod_j^m p(v_j, h_i|\theta)} \tag{7}$$

So for $v_j = 1$, the conditional probability is:

$$p(v_j|h,\theta) = \frac{\prod_i^n e^{-W_{ij}h_i - b_j}}{1 + \prod_i^n e^{-W_{ij}h_i - b_j}} \tag{8}$$

Again, since the nodes are independent and v can only take two values. The expectation terms can be simplified to:

$$E[v_j|h,\theta] = \frac{\prod_i^n e^{-W_{ij}h_i - b_j}}{1 + \prod_i^n e^{-W_{ij}h_i - b_j}} \tag{9}$$

$$= sigmoid(\sum_{i=1}^n W_{ij}h_i + b_j) \tag{10}$$

2.

The prove is the same as the first part. Just filp v and h:

$$p(h_i|v,\theta) = \frac{\prod_j^m e^{-W_{ij}v_j - c_i}}{1 + \prod_j^m e^{-W_{ij}v_j - c_i}} \tag{11}$$

$$\begin{aligned} E[h_i|v,\theta] &= \frac{\prod_j^m e^{-W_{ij}v_j - c_i}}{1 + \prod_j^m e^{-W_{ij}v_j - c_i}} && (12) \\ &= sigmoid(\sum_{j=1}^{m} W_{ij}v_j + c_i) && (13) \end{aligned}$$

For each training token:

$$log\ p(v|\theta) = log \sum_h e^{-E(v,h)} - logZ(\theta) \tag{14}$$

First let's find out the derivation of the first term in the RHS of eq.14:

$$\frac{-\sum_h e^{-E(v,h)} \frac{\partial E(v,h)}{\partial \theta}}{\sum_h e^{-E(v,h)}} = -\sum_h \frac{\partial E(v,h)}{\partial \theta} p(h|v) \tag{15}$$

Then for the second term:

$$\begin{aligned} \frac{1}{Z(\theta)} \frac{\partial Z(\theta)}{\partial \theta} &= \frac{-\sum_h \sum_v e^{-E(v,h)} \frac{\partial E(v,h)}{\partial \theta}}{\sum_h \sum_v e^{-E(v,h)}} && (16) \\ &= -\sum_h \sum_v \frac{\partial E(v,h)}{\partial \theta} p(v,h|\theta) && (17) \end{aligned}$$

So the derivative can be expressed as:

$$\begin{aligned} \frac{\partial \mathcal{L}(D|\theta)}{\partial \theta} &= \sum_h \sum_v \frac{\partial E(v,h)}{\partial \theta} p(v,h|\theta) - \sum_h \frac{\partial E(v,h)}{\partial \theta} p(h|v) && (18) \\ &= \mathbf{E}[\frac{\partial E(v,h)}{\partial \theta}|\theta] - \mathbf{E}[\frac{\partial E(v,h)}{\partial \theta}|v,\theta] && (19) \end{aligned}$$

3.

$$\frac{\partial E(v,h)}{\partial w_{ij}} = -h_i v_j \tag{20}$$

Plug the above expression into eq.19:

$$\begin{aligned} \frac{\partial \mathcal{L}(D|\theta)}{\partial w_{ij}} &= -\sum_h \sum_v v_j h_i p(v,h|\theta) + \sum_h v_j h_i p(h|v) && (21) \\ &= \mathbf{E}[v_j h_i|v,\theta] - \mathbf{E}[v_j h_i|\theta] && (22) \\ &= v_j sigmoid(\sum_{j=1}^{m} w_{ij}v_j + c_i) - \mathbf{E}[v_j h_i|\theta] && (23) \end{aligned}$$

for b:

$$\frac{\partial E(v,h)}{\partial b_j} = -v_j \tag{24}$$

Again plug it into eq.19:

$$\begin{aligned}
\frac{\partial \mathcal{L}(D|\theta)}{\partial b_j} &= -\sum_h \sum_v v_j p(v,h|\theta) + \sum_h v_j p(h|v) && (25)\\
&= \mathbf{E}[v_j|v,\theta] - \mathbf{E}[v_j|\theta] && (26)\\
&= v_j - \mathbf{E}[v_j|\theta] && (27)
\end{aligned}$$

For c:

$$\begin{aligned}
\frac{\partial \mathcal{L}(D|\theta)}{\partial c_i} &= -\sum_h \sum_v h_i p(v,h|\theta) + \sum_h h_i p(h|v) && (28)\\
&= \mathbf{E}[h_i|v,\theta] - \mathbf{E}[h_i|\theta] && (29)\\
&= sigmoid(\sum_{j=1}^{m} w_{ij} v_j + c_i) - \mathbf{E}[h_i|\theta] && (30)
\end{aligned}$$

4.

To compute the second term in (23), (27) and (30) require computing the normalization term $Z(\theta)$, which is intractable. Use Hinton approximation:

$$\begin{aligned}
\mathbf{E}[v_j h_i|\theta] &= \mathbf{E}[v_j|h]\mathbf{E}[h_i|v] && (31)\\
&= \mathbf{E}[v_j|\mathbf{E}[h|v,\theta]]\mathbf{E}[h_i|v] && (32)\\
&= \sum_{i=1}^{n} [w_{ij} sigmoid(\sum_{j=1}^{m} w_{ij} v_j + c_i)] sigmoid(\sum_{j=1}^{m} w_{ij} v_j + c_i) && (33)
\end{aligned}$$

for b:

$$\begin{aligned}
\mathbf{E}[v_j|\theta] &= \mathbf{E}[v_j|\mathbf{E}[h|v,\theta]] && (34)\\
&= \mathbf{E}[v_j|\mathbf{E}[h|v,\theta]]\mathbf{E}[h_i|v] && (35)\\
&= \sum_{i=1}^{n} [w_{ij} sigmoid(\sum_{j=1}^{m} w_{ij} v_j + c_i)] && (36)
\end{aligned}$$

for c:

$$\mathbf{E}[h_i|\theta] = sigmoid(\sum_{j=1}^{m} w_{ij} v_j + c_i) \tag{37}$$