Problem Set 4
ECE 685 Fall 2020

Name: Guillem Amat Castello

Problem 3: Bayesian Networks

3.1 Indicate whether the following items are True of False and explain why:

- $(H \perp\!\!\!\perp L)|P$: False. L also depends on G, which depends on H.
- $(H \perp\!\!\!\perp L)|P, G$: True. L only depends on P and G, and if they are already given, L is independent from H.
- $(P \perp\!\!\!\perp D)|G$: True. P depends on H and G. H is a root and G is already given, so D does not affect P.
- $(H \perp\!\!\!\perp D)|L$: False. H and D are dependent on the value of G, which depend on the value of L.

3.2 Write down the factorized form of the joint distribution $\mathbb{P}(H, D, P, G, L)$:

$\mathbb{P}(H, D, P, G, L) = P(H)P(D)P(G|H, D)P(P|H, G)P(L|G, P)$

3.3 Write down the factorized form of the joint distribution:

$\mathbb{P}(L = 1) = P(L = 1|P = A, G = A)P(P = A, G = A) + P(L = 1|P = A, G = B)P(P = A, G = B) + P(L = 1|P = B, G = A)P(P = B, G = A) + P(L = 1|P = B, G = B)P(P = B, G = B) = 0.8 * 0.4 + 0.3 * 0.6 + 0.1 * 0.3 + 0.2 * 0.1 = 0.55$

Problem 4: Restricted Boltzmann Machines

4.1 Show that the conditional distributions $p(x_k = 1|h)$ and $p(h_j = 1|x)$ are respectively given by $p(x_k = 1|h) = \sigma(c_k + h^T W_{.k})$ and $p(h_j = 1|x) = \sigma(b_j + W_{j.}x)$

We will do it for $p(h_j = 1|x)$, but we can reach the same results for $p(x_k = 1|h)$, using the same process. We can start with the conditional probability definition and then substituting the joint probability distribution in the 3rd step and the energy function in the last step:

$$\begin{aligned}
p(h_j = 1|x) &= \frac{p(x, h_j = 1)}{p(x)} \\
&= \frac{\sum_{h|h_k=1} p(x,h)}{\sum_h p(x,h)} \\
&= \frac{\sum_{h|h_k=1} exp[-E(x,h)]}{\sum_h exp[-E(x,h)]} \\
&= \frac{1}{1 + \frac{\sum_{h|h_k=1} exp[-E(x,h)]}{\sum_{h|h_k=0} exp[-E(x,h)]}} \\
&= \frac{1}{1 + \frac{\sum_{h|h_k=1} exp\left[\sum_{j,k} W_{j,k}h_j x_k + \sum_k c_k x_k + \sum_j b_j h_j)\right]}{\sum_{h|h_k=0} exp\left[\sum_{j,k} W_{j,k}h_j x_k + \sum_k c_k x_k + \sum_j b_j h_j)\right]}}
\end{aligned} \tag{1}$$

Finally, we can substitute the values of h in the last step of (1) to get:

$p(h_j = 1|x) = \frac{1}{1+e^{-b_j - \sum x_k W_{j,k}}} = \sigma(b_j + W_{j.}x)$

4.2 Negative Log-Likelihood(NLL) expression:

We start by developing the following expression:

$$\frac{\partial log p(x)}{\partial \theta} = \frac{\partial log \frac{\sum_h exp[-E(x,h)]}{\sum_{h,x} exp[-E(x,h)]}}{\partial \theta} = \frac{\partial log \sum_h exp[-E(x,h)]}{\partial \theta} - \frac{\partial log \sum_{h,x} exp[-E(x,h)]}{\partial \theta}$$

We now apply the chain rule twice to get the following expression:

$$\sum \left(\frac{exp[-E(x,h)]}{\sum exp[-E(x,h)]} * \frac{\partial[-E(x,h)]}{\partial \theta} \right) - \sum \left(\frac{exp[-E(x,h)]}{\partial log \sum_{h,x} exp[-E(x,h)]} * \frac{\partial[-E(x,h)]}{\partial \theta} \right)$$

Finally, we can substitute the joint and the hidden unit conditional distribution to get:

$$\sum p(h|x) * [-E(x,h)] - \sum p(x) \sum p(h|x) * (-E(x,h))$$

This proves that:

$$\frac{\partial log p(x)}{\partial \theta} = -\mathbb{E}_h \left[\frac{\partial E(x,h)}{\partial \theta} |x \right] + \mathbb{E}_{h,x} \left[\frac{\partial E(x,h)}{\partial \theta} \right]$$