CMPT 419 Graphical Models/ Recurrent Neural Networks

Echo Liu

Fall 2018

1 Question 1: Graphical Models

1. See Appendix section

2. $P(A, L, G, E, T) = P(T) * P(A|T) * P(E|L, G) * P(L)$

3. At Node T it is simply a Linear Gaussian $\frac{1}{\sqrt{2*\pi*\sigma^2}} * e^{\frac{-(x-\mu)^2}{2*\sigma^2}}$
   At Node A it is a sigmoidal distibution with the following form $\frac{1}{1+exp(g+\sum_{n=1}^{K} w_n*z_n)}$
   where the parents are T and E
   At Node E it is a Gaussian as follows, we use $L = 0$ to denote non-university and $L = 1$ to denote u and the same analogy applies for the current provincial government random variable G,
   $L(E|L = 0, G = 0) \sim\ N(A, B)$
   $L(E|L = 0, G = 1) \sim\ N(C, D)$
   $L(E|L = 1, G = 0) \sim\ N(E, F)$
   $L(E|L = 1, G = 1) \sim\ N(G, H)$
   As for the values of the respective parameters, it can be guessed that one would assign a larger

   μ for the Node T as it can be argued that there is more emphasis on the Tuition distribution when deciding whether or not to enroll in SFU. Furthermore, larger weighting
   will be given to the T parent in the Sigmoidal distribution as despite economy size, tuition becomes a larger deciding factor and the parameters for the Gaussian can be arbitrary.
   Some example values I'd use A=10, B=4, C=8,D=3, E=5, F=9, G=10, H=15.

4. We need to moralize our graph and realize that the only elements that are relevant are the intermediate parents at each of the nodes.

2 Question 2: KL Divergence

1. This is essentially the case when the two distributions are equal
   i.e $p(x) = q(x)$ As such when both distributions are equal the log inside the KL formula i.e $\frac{log(p(x))}{log(p(x))}$

2. No

3. We take the proof outlined in this Stats Stack Exchange post and change the "+" sign to a minus
sign and change it to a "-" sign and since our mu terms equal they also zero out plus our
0.5 terms zero out
$log(\frac{\sigma_2}{\sigma_1}) + \frac{\sigma_1^2}{2*\sigma_2^2} - log(\frac{\sigma_1}{\sigma_2}) - \frac{\sigma_2^2}{2*\sigma_1^2}$
$log(\frac{\sigma_2^2}{\sigma_1^2}) + \frac{\sigma_1^2}{2*\sigma_2^2} - \frac{\sigma_2^2}{2*\sigma_1^2}$
We then use the above equation and make the term inside the log our x

3 Question 3: Gated Recurrent Unit

1. When both are zero.

2. The hidden state at that time step will be also zero, i.e that GRU forgets everything.

4 Appendix