CS461 Homework 5

Due: Optional (Due Dec. 16 11:59pm)

1. [EM Algorithm] Using the four data points provided below, perform a single iteration of the Expectation-Maximization (EM) algorithm for a Gaussian Mixture Model (GMM): $f_X(x) = \pi_0 \cdot \mathcal{N}(\mu_0, \sigma_0^2) + \pi_1 \cdot \mathcal{N}(\mu_1, \sigma_1^2)$.

data num	x
d_1	2
d_2	1
d_3	-1
d_4	-2

1.1 Compute the log-likelihood for the parameters provided below. Assume that the four points are independent and identically distributed (i.i.d).

- $\pi_0(t) = \pi_1(t) = 1/2$
- $\mu_0(t) = -1, \mu_1(t) = 1$
- $\sigma_0^2(t) = \sigma_1^2(t) = 1.$

1.2 E-step: compute γ_{n0} **and** γ_{n1} **for all** $n = 1, 2, 3, 4$.

1.3 M-step: update the parameters $\pi_0(t+1), \pi_1(t+1), \mu_0(t+1), \mu_1(t+1), \sigma_0^2(t+1), \sigma_1^2(t+1)$.

1.4 Compute log-likelihood using the updated parameters and check the log-likelihood value increases.

2. [Exact vs. Approximate Inference] Given the Bayesian network below, compute P[Cloudy | Sprinkler = T, WetGrass = T] by using Variable Elimination and Gibbs Sampling.

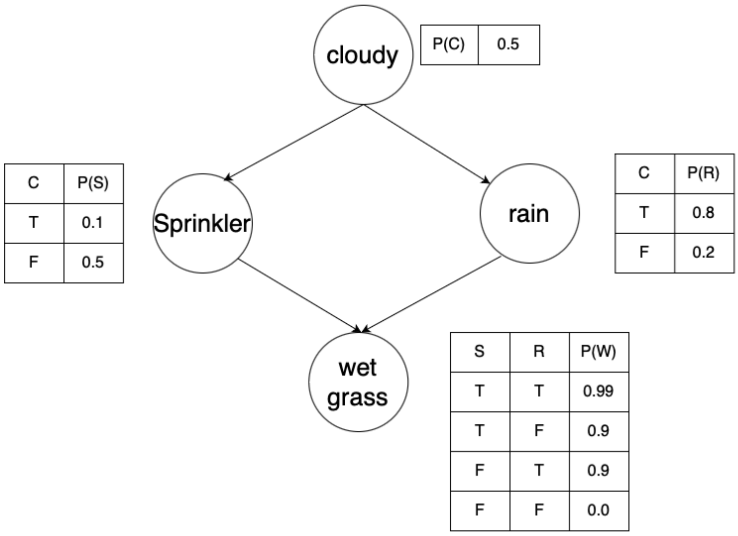


Figure 1: Bayesian Network

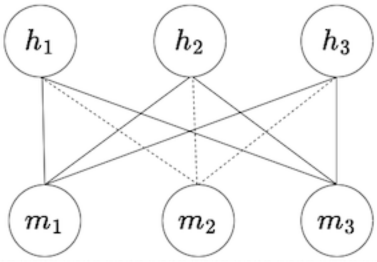
2.1 Compute the posterior by using Variable Elimination.

2.2 Estimate the posterior by using Gibbs Sampling.

3. [VAE Evidence Lower Bound (ELBO)] Derive the following inequality below.

$$\begin{aligned}
\log p_\theta(x_i) &= \log \sum_z p_\theta(x_i, z) \\
&= \log \sum_z \frac{p_\theta(x_i, z) q_\phi(z|x_i)}{q_\phi(z|x_i)} \\
&\geq E_{q_\phi(z|x_i)}[\log p_\theta(x_i, z) - \log q_\phi(z|x_i)] \\
&= -D_{KL} q_\phi(z|x_i) || p_\theta(z) + E_{q_\phi(z|x_i)}[\log p_\theta(x_i|z)]
\end{aligned}$$

4. [RBM Movie Recommendation System] It is known that a part of Netflix's recommendation system utilizes a Restricted Boltzmann Machine (RBM). Suppose you are given an RBM and energy function as below and the preference of a user for movie1 (m_1) and movie3 (m_3), but no information is given for movie2 (m_2). Predict if the user likes movie2 or not. For the prediction, you will need a sampling process. Assume the sampling process is deterministic; $x = 1$ if $P(x) > 1/2$ and $x = 0$ otherwise.



$$\mathcal{E}(v, h) = -v^t W h - b^t v - c^t h$$

W			
	-1	1	0
	0	1	2
	-1	-2	3

b	
	1
	-1
	1

c	
	-2
	0
	1

Figure 2: RBM and Energy Function of Recommendation System

4.1 In this recommendation system, the RBM needs to operate on bipolar coding $(+1, -1)$ instead of using binary coding 1 and 0. Why is this necessary?

4.2 In the class, we studied the conditional probability of an RBM using binary coding 1 and 0. The formulation is shown below. How would you modify the formulation for the RBM using bipolar coding?

$$P(h_1 = 1 | m_1, m_2, m_3) = \frac{1}{1 + \exp^{-m^t W[:,1] - c_1}} = \sigma(m^t W[:,1] + c_1)$$

$$P(m_1 = 1 | h_1, h_2, h_3) = \frac{1}{1 + \exp^{-W[1,:]h - d_1}} = \sigma(W[1,:]h + b_1)$$

4.3 Suppose the user liked m_1 but disliked m_3. Based on the user's preference, predict the preference for m_2.