Korea Advanced Institute of Science and Technology
School of Electrical Engineering
EE488 Intro. to Machine Learning, Spring 2018

Issued: June 2, 2018
Due: June 12, 2018

Assignment IV

Policy

Group study is encouraged; however, assignment that you hand-in must be of your own work. Anyone suspected of copying others will be penalized. The homework will take considerable amount of time so start early.

1. **EM:** Assume that there are only two grades for qualifying the quality of a watermelon: high and low grade. In addition, assume that there are only two stores $\{S_i\}_{i=1}^2$ selling the watermelons. Each store has different proportion of high-grade(H) and low-grade (L) watermelons. S_1 has $\alpha\%$ high-grade watermelons while S_2 has $\beta\%$ high-grade watermelons. S_1 is γ times more accessible than S_2. Seven watermelons are bought by a watermelon lover $\mathbf{O} = \{o_1, o_2, \ldots, o_7\} = \{H, H, H, H, L, L, H\}$ who dies immediately upon purchasing the watermelons. It is not known where each watermelon was purchased.

   (i) What is the probability that the second watermelon $o_2 = \mathtt{H}$ was bought from $q_2 = S_1$, in otherwords, what is $P(q_2 = S_1 | o_2 = \mathtt{H}, \theta)$. The watermelons are not necessarily purchased from the same store.

   (ii) Define the auxiliary function $Q(\theta, \bar{\theta})$ where $\theta = \{\alpha, \beta, \gamma\}$.

   (iii) Derive/define the EM steps in estimating θ. We make no Markov assumptions.

2. **EM:** The log likelihood $p(\mathbf{X}|\theta)$ can be represented as the sum of two terms. The first term is given as $\sum_{\text{all possible } \mathbf{H}} q(\mathbf{H}) \ln \left\{ \frac{p(\mathbf{H}|\mathbf{X},\theta)}{q(\mathbf{H})} \right\}$. This term is often referred to as the Kullback-Leibler (KL) divergence between $q(\mathbf{H})$ and $p(\mathbf{H}|\mathbf{X}, \theta)$ and is denoted as $\mathrm{KL}(q||p)$. Here $q(\cdot)$ represents a probability. The KL divergence has two important properties: (1) it is asymmetric $\mathrm{KL}(q||p) \neq \mathrm{KL}(p||q)$ and (2) $\mathrm{KL}(q||p) \geq 0$. It measures how $q(\cdot)$ differs from $p(\cdot)$ and when $q(\cdot) = p(\cdot)$ then its value is 0.

   (i) Express the second term using $q(\cdot)$. It should be realized that for a fixed θ and given data $\mathbf{X}$, the sum of the first and second term is constant irrespective of how $q(\cdot)$ and $\mathbf{H}$ are defined.

   (ii) The above representation of the log likelihood can be used to explain the EM algorithm. Derive the EM algorithm by iteratively fixing the θ to θ^{old} and estimating $q(\mathbf{H})$ then fixing $q(\mathbf{H})$ to its estimated value and estimating for θ and setting it to θ^{old}.

3. **EM:** Given three samples $D = \{(1,3), (4,5), (2,*)\}$ which are sampled from a two dimensional distribution $p_\theta(x_1, x_2) = p_{\theta_1}(x_1) p_{\theta_2}(x_2)$ where

$$p_{\theta_1}(x_1) = \begin{cases} \frac{1}{\theta_1} \exp -x_1/\theta_1 & \text{if } x_1 \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad p_{\theta_2}(x_2) = \begin{cases} \frac{1}{\theta_2} & \text{if } \theta_2 < 0 \\ 0 & \text{otherwise} \end{cases}$$

   (i) What can you say about θ_2 from D?

(ii) Determine the auxiliary function $Q(\theta, \bar{\theta})$ and use initial point as $\theta^{(0)} = (3, 6)$ determine expectation step and the maximization step.

4. **HMM**: Casino488 has a game where the player and the casino each take turn throwing a die. The player wins the bet when he/she has a higher number on the roll. Player bets 1 dollars for each roll and takes 2 dollars when he/she wins. Unknown to the player, the casino uses two dice: a fair and load die. For the fair die, all numbers between 1 through 6 have equal probability of appearing while for the loaded die, $p(1) = p(2) = \ldots = p(5) = 1/12$ $p(6) = 7/12$. The casino will switch die with probability of 1/4 irrespective of the die. The probability that the casino starting with a fair die is 1/3.

   (i) Model the casino's game operation as a 2 state HMM.

   (ii) What is p(casino uses fair die at second roll)?

   (iii) What is p(casino rolls 4 on second roll)?

   (iv) Casino rolls $\{6, 6, 2, 6, 6, 2, 4, 5\}$ what is the most probable sequence of fair and loaded die?

   (v) What is p(casino rolls 3 on 100th roll)?

   (vi) What is p(use fair die on second roll|6 on the second roll)

   (vii) Compute $\arg\max_{s_1,s_2,s_3,s_4} p(q_1 = s_1, q_2 = s_2, q_3 = s_3, q_4 = s_4 | o_1 = o_2 = o_3 = o_4 = 6)$ where q_t and o_t are the state and observation variables respectively. Also $s_t \in \{\text{fair}, \text{loaded}\}$.