**CS-419m: Practice question set 5**

1. **With answers** Consider the following eight points with $d = 3$ binary attributes. Assume that we measure distance between two points $\mathbf{x}^i$ and $\mathbf{x}^j$ as $d(\mathbf{x}^i, \mathbf{x}^j) = \sum_{k=1}^{d} |x_k^i - x_k^j|$.

| | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|
| $\mathbf{x}^0$ | 0 | 0 | 0 |
| $\mathbf{x}^1$ | 0 | 0 | 1 |
| $\mathbf{x}^2$ | 0 | 1 | 0 |
| $\mathbf{x}^3$ | 0 | 1 | 1 |
| $\mathbf{x}^4$ | 1 | 0 | 0 |
| $\mathbf{x}^5$ | 1 | 0 | 1 |
| $\mathbf{x}^6$ | 1 | 1 | 0 |
| $\mathbf{x}^7$ | 1 | 1 | 1 |

(a) Show the clustering at the end of an iteration of K-means where the number of clusters $K = 2$ and the starting centroids are at $m_1 = [1\ 0\ 0]$ and $m_2 = [0\ 1\ 1]$

- Members of cluster 1:

  $x_0, x_4, x_5, x_6$

- Members of cluster 2:

  $x_1, x_2, x_3, x_7$ ..1

(b) Now we will apply the EM algorithm on this data. Assume the probability of an instance in each cluster follows an independent Bernoulli distribution that is, $\Pr(\mathbf{x}|k) = B(p_{k1}, x_1)B(p_{k2}, x_2)B(p_{k3}, x_3)$ where $B(p, x) = p^x(1-p)^{1-x}$ is the probability of $x$ under a Bernoulli distribution with parameter $p$. For example, $\Pr(\mathbf{x}^3|k) = (1 - p_{k1})p_{k2}p_{k3}$.

i. Say at iteration $t$ $[p_{11}^t \quad p_{12}^t \quad p_{13}^t]$ is $[0.5\ 0.5\ 0.1]$ and $[p_{21}^t \quad p_{22}^t \quad p_{23}^t]$ is $[0.5\ 0.5\ 1]$ and $\pi_1^t = \pi_2^t = 1/2$. Write the $z_{ik}^t$ values for each instance in this table. [Note: $z_{ik} = \Pr(k|\mathbf{x}^i, \theta^t)$]

| | $z_{i1^t}$ | $z_{i2}^t$ |
|---|---|---|
| $\mathbf{x}^0$ | 1 | 0 |
| $\mathbf{x}^1$ | $\frac{1}{1.1}$ | $\frac{0.1}{1.1}$ |
| $\mathbf{x}^2$ | 1 | 0 |
| $\mathbf{x}^3$ | $\frac{1}{1.1}$ | $\frac{0.1}{1.1}$ |
| $\mathbf{x}^4$ | 1 | 0 |
| $\mathbf{x}^5$ | $\frac{1}{1.1}$ | $\frac{0.1}{1.1}$ |
| $\mathbf{x}^6$ | 1 | 0 |
| $\mathbf{x}^7$ | $\frac{1}{1.1}$ | $\frac{0.1}{1.1}$ |

ii. What will be the value of $p_{13}^{t+1}$ and $p_{23}^{t+1}$ at the end of the $M$ step of this iteration.
$p_{23}^{t+1} = 1$
$p_{13}^{t+1} = \frac{4*\frac{0.1}{1.1}}{4*\frac{0.1}{1.1}+4*1} = \frac{0.1}{1.2} < 0.1$ ..2

iii. What will be the final converged values of the parameters at the end of all EM iterations. [Justify briefly]
`With each iteration,` $p_{13}^{t+1}$ `reduces and finally converges to 0.`
$p_{23}^{t+1} = 1$ ..2

iv. Do you see a nice description of the two kinds of clusters that the algorithm produces [Hint: The instances denote binary representation of numbers] `All odd numbers go to one cluster, and even numbers go to another cluster.` ..1

v. What is another set of parameters where the EM algorithm achieves local optimality. [Justify briefly].
$p_1^0$ `= [0.5 0.5 0.5]`
$p_2^0$ `= [0.5 0.5 0.5]` ..2

2. **Without answers** Consider the following estimation problem. Your friend has two coins with probability of coming up heads for coin 1 $= p_1$ and coin 2 $= p_2$. He performs this experiment four times: choose coin 1 with probability $\lambda$ or coin 2 with probability $1 - \lambda$ and toss the chosen coin three times. Let the observed outcome of coin tosses be $Y = \{\{HHH\}, \{TTT\}, \{HHH\}, \{TTT\}\}$. Let $y_i$ for $i = 1, 2, 3, 4$ denote the coin chosen in each experiment.

(a) First, consider the case where the value of $y_i$ is known and is observed to be $(z_1, z_2, z_3, z_4) = (1, 2, 1, 1)$. Write down the maximum likelihood estimation of the parameters.

i. $\lambda$
$= 3/4$ ..1

ii. $p_1$
$= 6/9 = 2/3$ ..1

iii. $p_2$
$= 3/3 = 1$ ..1

(b) Now consider the case where $y_i$ is not known. We will use EM to estimate the parameters $(\lambda, p_1, p_2)$. Let $(\lambda^t, p_1^t, p_2^t)$ denote the value of the parameter at the beginning of the $t$-th iteration. For the **E** step of EM, express in full expanded form the value of $z_i = \Pr(y_i = 1|h_i)$ in terms of these and the absolute counts $h_i$ of heads in the $i$-th experiment in $Y$. ..3

$z_i = \frac{\lambda p_1^{h_i}(1-p_1)^{3-h_i}}{\lambda p_1^{h_i}(1-p_1)^{3-h_i}+(1-\lambda)p_2^{h_i}(1-p_2)^{3-h_i}}$

(c) For the **M** step, express the updated parameter values, in terms of expected values above and the absolute counts of Hs and Ts in each elemnt of $Y$.

i. $\lambda$
$= \sum_i z_i/4$ ..1

ii. $p_1$
$= \frac{\sum_i z_i h_i}{\sum_i 3z_i}$ ..2

iii. $p_2$
$= \frac{\sum_i (1-z_i)h_i}{\sum_i 3(1-z_i)}$ ..1

(d) Identify an assignment of $(\lambda^*, p_1^*, p_2^*)$ at which EM cannot make progress but which is not globally optimal. You need to provide the value of data likelihood at this assignment and another assignment for which likelihood is higher and show that the **E** and **M** steps cause no change at $(\lambda^*, p_1^*, p_2^*)$. (Hint: try to guess the optimal parameter values from the $Y$s) ..4 `Optimal assignment is` $(\lambda^*, p_1^*, p_2^*) = (0.5, 1, 0)$ `At this the data likelihood is` $4 \log 0.5$. `A locally optimal assignment is` $(\lambda, p_1, p_2) = (0.5, 0.5, 0.5)$ `for which likelihood is` $4 \log(0.5 * (0.5^3 + 0.5^3))$ `which is less than optimal. For this assignment, the value of` $z_i$ `is 0.5, and the parameter values stay at (0.5, 0.5, 0.5) in the M-step.`