Q1-1: You are presented with a dataset that has hidden/missing variables that influences your data. You are asked to use Expectation Maximization algorithm to best capture the data. How would you define the E and M in Expectation Maximization?

1. Estimate the Missing/Latent Variables in the Dataset, Maximize the likelihood over the parameters in the model
2. Estimate the number of Missing/Latent Variables in the Dataset, Maximize the likelihood over the parameters in the model
3. Estimate likelihood over the parameters in the model, Maximize the number of Missing/Latent Variables in the Dataset
4. Estimate the likelihood over the parameters in the model, Maximize the number of parameters in the model
Q1-1: You are presented with a dataset that has hidden/missing variables that influences your data. You are asked to use Expectation Maximization algorithm to best capture the data. How would you define the $E$ and $M$ in Expectation Maximization?

1. Estimate the Missing/Latent Variables in the Dataset, Maximize the likelihood over the parameters in the model
2. Estimate the number of Missing/Latent Variables in the Dataset, Maximize the likelihood over the parameters in the model
3. Estimate likelihood over the parameters in the model, Maximize the number of Missing/Latent Variables in the Dataset
4. Estimate the likelihood over the parameters in the model, Maximize the number of parameters in the model
Q1-2: Select the correct statement.

A. The EM algorithm is guaranteed to converge but may not reach a global optimum.
B. The objective function optimized by the EM algorithm can also be optimized by a gradient descent algorithm which will find the global optimal solution, whereas EM finds its solution more quickly but may return only a locally optimal solution.

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.
Q1-2: Select the correct statement.

A. The EM algorithm is guaranteed to converge but may not reach a global optimum.
B. The objective function optimized by the EM algorithm can also be optimized by a gradient descent algorithm which will find the global optimal solution, whereas EM finds its solution more quickly but may return only a locally optimal solution.

1. Both the statements are TRUE.
2. Statement A is TRUE, but statement B is FALSE.
3. Statement A is FALSE, but statement B is TRUE.
4. Both the statements are FALSE.

For the second statement:
The only false part is that the gradient descent algorithm will find the global optimal solution. Gradient descent can also get stuck in a local optima.
Q2-1: Select the correct statement.

A. The Chow-Liu algorithm not necessarily always choose edges from a complete graph.
B. The algorithm tries to find a minimum spanning tree of a graph to minimize the negative log-likelihood of training data.
C. Edge directions can be assigned randomly in the Chow-Liu algorithm.

1. True, True, True
2. False, False, True
3. True, False, True
4. False, False, False
Q2-1: Select the correct statement.

A. The Chow-Liu algorithm not necessarily always choose edges from a complete graph.
B. The algorithm tries to find a minimum spanning tree of a graph to minimize the negative log-likelihood of training data.
C. Edge directions can be assigned randomly in the Chow-Liu algorithm.

1. True, True, True
2. False, False, True
3. True, False, True
4. False, False, False

1. The Chow-Liu algorithm always have a complete graph.
2. The algorithm tries to find a maximum spanning tree of a graph to minimize the negative log-likelihood of training data.
3. Any directions for edges: Once we pick a node, and edges going away from this node, so that it remains a tree.
Q2-2: Which of the following can NOT be the sequence of edges added, in that order, to a maximum spanning tree using Kruskal’s algorithm?

1. (c - f), (a - c), (e - f), (b - d), (b - c)
2. (c - f), (a - c), (e - f), (c - e), (b - d)
3. (c - f), (a - c), (e - f), (b - d), (d - e)
4. All of the above are valid.
Q2-2: Which of the following can NOT be the sequence of edges added, in that order, to a maximum spanning tree using Kruskal’s algorithm?

1. (c - f), (a - c), (e - f), (b - d), (b - c)
2. (c - f), (a - c), (e - f), (c - e), (b - d)
3. (c - f), (a - c), (e - f), (b - d), (d - e)
4. All of the above are valid.

(c - f), (a - c), (e - f), (c - e), (b - d) form a cycle.
Q3-1: Select the correct statement.

A. *Sparse Candidate Algorithm (SCA) is an iterative algorithm.*
B. *SCA consists of 2 parts: Restrict Phase and Maximize Phase.*
C. *SCA will always lead to a global optimal solution.*

1. True, True, True
2. True, False, True
3. True, True, False
4. False, True, False
Q3-1: Select the correct statement.

A. Sparse Candidate Algorithm (SCA) is an iterative algorithm.
B. SCA consists of 2 parts: Restrict Phase and Maximize Phase.
C. SCA will always lead to a global optimal solution.

1. True, True, True
2. True, False, True
3. True, True, False
4. False, True, False

SCA can lead to sub-optimal solution.
Q3-2: Recall for Bernoulli distribution: Let $X \sim \text{Bern}(\theta)$, $x \in \{0, 1\}$, $0 < \theta < 1$. Then, $p_\theta(x) = \theta^x(1 - \theta)^{1-x}$ and $E[X] = \theta$. Consider two bernoulli distributions $p_{\theta_1}(X)$ and $p_{\theta_2}(X)$. Calculate the KL divergence: $KL(p_{\theta_1}(X) \| p_{\theta_2}(X))$.

1. $\theta_1 \log[\theta_1/\theta_2] + (1 - \theta_1) \log[(1 - \theta_1)/(1 - \theta_2)]$

2. $\theta_2 \log[\theta_1/\theta_2] + (1 - \theta_2) \log[(1 - \theta_1)/(1 - \theta_2)]$

3. $(1 - \theta_1) \log[\theta_1/\theta_2] + \theta_1 \log[(1 - \theta_1)/(1 - \theta_2)]$

4. $(1 - \theta_2) \log[\theta_1/\theta_2] + \theta_2 \log[(1 - \theta_1)/(1 - \theta_2)]$

$$D_{KL}(P(X) \| Q(X)) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$$
Q3-2: Recall for Bernoulli distribution: Let $X \sim \text{Bern}(\theta)$, $x \in \{0, 1\}$, $0 < \theta < 1$. Then, $p_\theta(x) = \theta^x(1 - \theta)^{1-x}$ and $E[X] = \theta$. Consider two bernoulli distributions $p_{\theta_1}(X)$ and $p_{\theta_2}(X)$. Calculate the KL divergence: $KL(p_{\theta_1}(X) \parallel p_{\theta_2}(X))$.

1. $\theta_1 \log[\theta_1/\theta_2] + (1 - \theta_1) \log[(1 - \theta_1)/(1 - \theta_2)]$

2. $\theta_2 \log[\theta_1/\theta_2] + (1 - \theta_2) \log[(1 - \theta_1)/(1 - \theta_2)]$

3. $(1 - \theta_1) \log[\theta_1/\theta_2] + \theta_1 \log[(1 - \theta_1)/(1 - \theta_2)]$

4. $(1 - \theta_2) \log[\theta_1/\theta_2] + \theta_2 \log[(1 - \theta_1)/(1 - \theta_2)]$

Log-Likelihood Ratio (LLR)

$= \log[p_{\theta_1}(X)/p_{\theta_2}(X)]$

$= \log[\theta_1^x(1 - \theta_1)^{1-x} / \theta_2^x(1 - \theta_2)^{1-x}]$

$= X \log[\theta_1/\theta_2] + (1 - X) \log[(1 - \theta_1)/(1 - \theta_2)]$

$KL(p_{\theta_1}(x) \parallel p_{\theta_2}(x)) = E_{\theta_1}(\text{LLR})$

$= E_{\theta_1}[X] \log[\theta_1/\theta_2] + (1 - E_{\theta_1}[X]) \log[(1 - \theta_1)/(1 - \theta_2)]$

$= \theta_1 \log[\theta_1/\theta_2] + (1 - \theta_1) \log[(1 - \theta_1)/(1 - \theta_2)]$