1.

(A) The covariance between $X_3$ and $X_4$ is given by the $(3, 4)$-th element of the covariance matrix $\Sigma$, i.e., $\Sigma_{3,4}$.
Given the covariance matrix:

$$\Sigma = \begin{pmatrix} 0.71 & -0.43 & 0.43 & 0 \\ -0.43 & 0.46 & -0.26 & 0 \\ 0.43 & -0.26 & 0.46 & 0 \\ 0 & 0 & 0 & 0.2 \end{pmatrix}$$

We see that $\Sigma_{3,4} = 0$, indicating that $X_3$ and $X_4$ are not correlated.

(B) Given the precision matrix $Q$,

$$Q = \begin{pmatrix} 5 & 3 & -3 & 0 \\ 3 & 5 & 0 & 0 \\ -3 & 0 & 5 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}$$

we can see that $Q_{3,4} = Q_{4,3} = 0$. This implies that $X_3$ and $X_4$ are conditionally independent given $X_1$ and $X_2$. Conditional independence implies that the conditional covariance $\mathbf{cov}(X_3, X_4 | X_1, X_2)$ is zero. Therefore, $X_3$ and $X_4$ are not conditionally correlated given $X_1$ and $X_2$.

(C) Given the precision matrix $Q$,

$$Q = \begin{bmatrix} 5 & 3 & -3 & 0 \\ 3 & 5 & 0 & 0 \\ -3 & 0 & 5 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

The non-zero entries in the second row (corresponding to $X_2$) are $Q_{1,2}$ and $Q_{2,2}$. In the context of Gaussian graphical models, we don't consider the variable itself as part of its Markov blanket, so we only include $X_1$. Therefore, the Markov blanket of $X_2$, denoted $X_{M_2}$, is $\{X_1\}$.

(D) We know that $Y = [Y_1, Y_2]^\top$ where $Y_1 = X_1 + X_4$ and $Y_2 = X_2 - X_4$. We can write this as a linear transformation $Y = AX$, where

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

The covariance of $Y$ is then $Cov(Y) = ACov(X)A^T = A\Sigma A^T$. So,

$$Cov(Y) = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0.71 & -0.43 & 0.43 & 0 \\ -0.43 & 0.46 & -0.26 & 0 \\ 0.43 & -0.26 & 0.46 & 0 \\ 0 & 0 & 0 & 0.2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & -1 \end{bmatrix}$$

After performing the matrix multiplication, we get

$$Cov(Y) = \begin{bmatrix} 0.91 & -0.63 \\ -0.63 & 0.66 \end{bmatrix}$$

Therefore, the covariance matrix of $Y$ is $\begin{bmatrix} 0.91 & -0.63 \\ -0.63 & 0.66 \end{bmatrix}$.

2.

(A) The EM algorithm includes two steps: the E-step and the M-step.
In the E-step, we calculate the responsibilities. The responsibility of a data point $x^{(i)}$ generated by a Gaussian component $j$ with mean $\mu_j$ is:

$$w_{ij} = \frac{\exp(-0.5 * (x^{(i)} - \mu_j)^2)}{\exp(-0.5 * (x^{(i)} - \mu_1)^2) + \exp(-0.5 * (x^{(i)} - \mu_2)^2)}$$

We use this formula to compute the responsibilities for our given data points $x^{(1)} = -1$ and $x^{(2)} = 1$ and the initial Gaussian means $\mu_1 = -2$ and $\mu_2 = 2$:

$$w_{11} = \frac{\exp(-0.5 * (-1 - (-2))^2)}{\exp(-0.5 * (-1 - (-2))^2) + \exp(-0.5 * (-1 - 2)^2)} = 0.98,$$

$$w_{12} = 1 - w_{11} = 0.02,$$

$$w_{21} = \frac{\exp(-0.5 * (1 - (-2))^2)}{\exp(-0.5 * (1 - (-2))^2) + \exp(-0.5 * (1 - 2)^2)} = 0.02,$$

$$w_{22} = 1 - w_{21} = 0.98.$$

In the M-step, we update the parameters using the computed responsibilities:

$$\mu_j^{new} = \frac{\sum_{i=1}^{2} w_{ij} x^{(i)}}{\sum_{i=1}^{2} w_{ij}}$$

This results in the following updated parameters:

$$\mu_1^{new} = \frac{0.98(-1) + 0.02(1)}{0.98 + 0.02} = -1,$$

$$\mu_2^{new} = \frac{0.02(-1) + 0.98(1)}{0.02 + 0.98} = 1.$$

Therefore, after the first iteration of the EM algorithm, the estimated parameters are $\mu_1 = -1$ and $\mu_2 = 1$.

(B) As explained above, the means $\mu_1$ and $\mu_2$ have converged to the data points $-1$ and $1$ respectively, meaning the EM algorithm has reached a solution where each Gaussian component corresponds to a single data point. As such, it will remain in this state in further iterations since the responsibilities and therefore the means will not change. Hence, the EM algorithm will indeed eventually converge to $\mu_1 = -1$ and $\mu_2 = 1$ when initialized with $\mu_1 = -2$ and $\mu_2 = 2$.

(C) If we initialize both means to $2$, then $\mu_1 = \mu_2 = 2$. In this case, the responsibilities $w_{ij}$ for each data point and each Gaussian component are equal, and both data points contribute equally to the update of both means. Thus, the updated means will be the same, equal to the average of the data points, which is $0$. As such, the fixed point of the EM algorithm when initialized from $\mu_1 = \mu_2 = 2$ is $\mu_1 = \mu_2 = 0$.

(D) With K-means, each data point is assigned to the closest mean. The data point $-1$ gets assigned to $\mu_1 = -2$ and $1$ gets assigned to $\mu_2 = 2$. In the update step, each mean is updated to the average of the data points assigned to it. Hence, $\mu_1$ gets updated to $-1$ and $\mu_2$ gets updated to $1$. Since there is no change in the assignment of data points to the means in the subsequent iterations, the algorithm has reached its fixed point, which is $\mu_1 = -1$, $\mu_2 = 1$.