Q1

1. For $SSE = \sum_{i=1}^{n} \| x_i - c_{z_i} \|_2^2$, it's minimized when its derivative with respect to c_j is equal to zero if we update the cluster means.

$$\frac{\partial SSE}{\partial c_j} = \frac{\partial}{\partial c_j} \sum_{j=1}^{k} (x_j - c_j)^2$$

$$= 2\sum_{j=1}^{k} (x_j - c_j) = 0$$

then, we can infer that $c_j = \frac{1}{\text{\# points assigned to cluster}_j} \sum x_j$.

So, the best c_j happens when it's equal to the mean of data points assigned to current cluster and updating the SSE function will either remain unchanged or move toward local minima with respect to c_j.

Also, when we update the cluster assignment, we're reducing l_2-norm distance between each data point and cluster centers which will decrease SSE.

At last, optimizations of both cluster mean and assignments reach the best value and the algorithm converges.