Problem 1

(a) We prove (10.12) for an arbitrarily given group $C_k$. suppose there are $n$ observations, $x_1, x_2 \cdots x_n$ in $C_k$. Then (10.12) becomes:

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{p}(x_{ij}-x_{kj})^2 = 2\sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij}-\bar{x}_j)^2$$

where $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n}x_{ij}$

$$\text{Left side} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{p}(x_{ij}-\bar{x}_j+\bar{x}_j-x_{kj})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{p}\left[(x_{ij}-\bar{x}_j)^2+(x_{kj}-\bar{x}_j)^2\right] - \frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{p}(x_{ij}-\bar{x}_j)(x_{kj}-\bar{x}_j)$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n}\sum_{j=1}^{p}n(x_{ij}-\bar{x}_j)^2+\sum_{k=1}^{n}\sum_{j=1}^{p}n(x_{kj}-\bar{x}_j)^2\right] - \frac{2}{n}\sum_{j=1}^{p}\left[\sum_{i=1}^{n}x_{ij}-n\bar{x}_j\right]\left[\sum_{k=1}^{n}x_{kj}-n\bar{x}_j\right]$$

$$= \frac{1}{n}\cdot 2n\cdot\sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij}-\bar{x}_j)^2 \qquad - \frac{2}{n}\sum_{j=1}^{p}\cdot 0\cdot 0$$

$$= 2\sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij}-\bar{x}_j)^2$$

$=$ right side

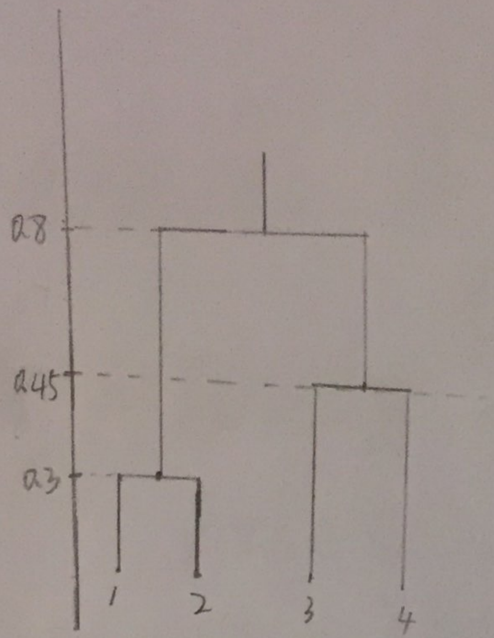
(b) For any given observation $x$, suppose it is originally assigned to group $i$, then in the next iteration, it is assigned to group $j$, this can only happen when:

$$\sum_{k=1}^{p}(x_k-\bar{x}_{jk})^2 \le \sum_{k=1}^{p}(x_k-\bar{x}_{ik})^2$$ because $j$'s centroid $\bar{x}_j=(\bar{x}_{j1}\cdots\bar{x}_{jp})'$ is now the cloest to $x=(x_1\cdots x_p)'$ according to algorithm 10.1 step (b)

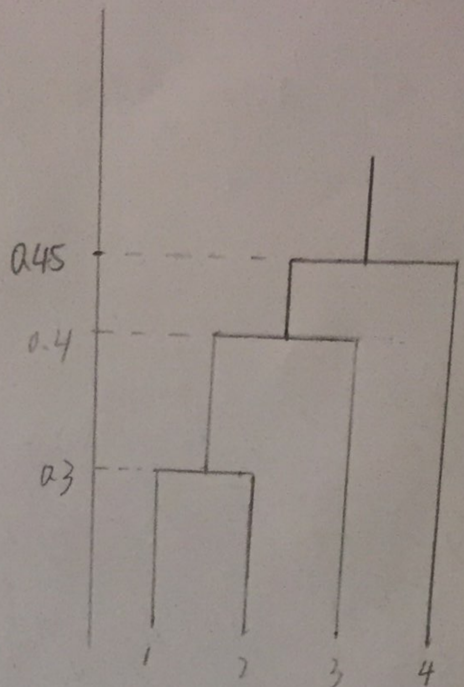
This holds true for all observations. Therefore according to 10.12, algorithm 10.1 decreases 10.11 at each iteration.

Problem 2

(a)




(b)




(c) 1,2 in cluster 1
3.4 in cluster 2

(d) 1,2,3 in cluster 1
4 in cluster 2

(e)