1. Proof:

(a) To prove

$$\frac{1}{|C_k|}\sum_{i,i'\in C_k}\sum_{j=1}^{p}(x_{ij}-x_{i'j})^2 = 2\sum_{i\in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})^2.$$

First we clarify the notation:

$$\sum_{i,i'\in C_k} = \sum_{i\in C_k}\sum_{i'\in C_k}.$$

This means that for each C_k, we are examining $|C_k|^2$ pairs, NOT $\binom{|C_k|}{2}$ pairs.

We do the following transformation to the lefthand side:

$$\frac{1}{|C_k|}\sum_{i,i'\in C_k}\sum_{j=1}^{p}(x_{ij}-x_{i'j})^2 = \frac{1}{|C_k|}\sum_{i,i'\in C_k}\sum_{j=1}^{p}\left[(x_{ij}-\bar{x}_{kj})-(x_{i'j}-\bar{x}_{kj})\right]^2$$

$$= \frac{1}{|C_k|}\sum_{i,i'\in C_k}\sum_{j=1}^{p}\left[(x_{ij}-\bar{x}_{kj})^2 - 2(x_{ij}-\bar{x}_{kj})(x_{i'j}-\bar{x}_{kj}) + (x_{i'j}-\bar{x}_{kj})^2\right]$$

$$= \frac{1}{|C_k|}\sum_{i,i'\in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})^2 + \frac{1}{|C_k|}\sum_{i,i'\in C_k}\sum_{j=1}^{p}(x_{i'j}-\bar{x}_{kj})^2$$

$$- \frac{2}{|C_k|}\sum_{i,i'\in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})(x_{i'j}-\bar{x}_{kj})$$

we made change here

$$= \frac{1}{|C_k|}\sum_{i\in C_k}\sum_{i'\in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})^2 + \frac{1}{|C_k|}\sum_{i\in C_k}\sum_{i'\in C_k}\sum_{j=1}^{p}(x_{i'j}-\bar{x}_{kj})^2$$

$$- \frac{2}{|C_k|}\sum_{i\in C_k}\sum_{i'\in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})(x_{i'j}-\bar{x}_{kj})$$

$$= \frac{|C_k|}{|C_k|}\sum_{i\in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})^2 + \frac{|C_k|}{|C_k|}\sum_{i'\in C_k}\sum_{j=1}^{p}(x_{i'j}-\bar{x}_{kj})^2$$

$$- \frac{2}{|C_k|}\sum_{i\in C_k}\sum_{i'\in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})(x_{i'j}-\bar{x}_{kj})$$

$$= 2\sum_{i\in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})^2 - \frac{2}{|C_k|}\sum_{i\in C_k}\sum_{i'\in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})(x_{i'j}-\bar{x}_{kj})$$

Next, we show the second half is zero:

$$\frac{2}{|C_k|}\sum_{i\in C_k}\sum_{i'\in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})(x_{i'j}-\bar{x}_{kj})$$

$$=2\sum_{j=1}^{p}\sum_{i\in C_k}\sum_{i'\in C_k}\frac{1}{|C_k|}(x_{ij}-\bar{x}_{kj})(x_{i'j}-\bar{x}_{kj})$$

$$=2\sum_{j=1}^{p}\sum_{i\in C_k}\left[(x_{ij}-\bar{x}_{kj})\left(\sum_{i'\in C_k}\frac{1}{|C_k|}(x_{i'j}-\bar{x}_{kj})\right)\right]$$

$$=2\sum_{j=1}^{p}\sum_{i\in C_k}\left[(x_{ij}-\bar{x}_{kj})\left(\frac{\sum_{i'\in C_k}x_{i'j}}{|C_k|}-\frac{|C_k|}{|C_k|}\bar{x}_{kj}\right)\right]$$

$$=2\sum_{j=1}^{p}\sum_{i\in C_k}\left[(x_{ij}-\bar{x}_{kj})(\bar{x}_{kj}-\bar{x}_{kj})\right] = 0$$

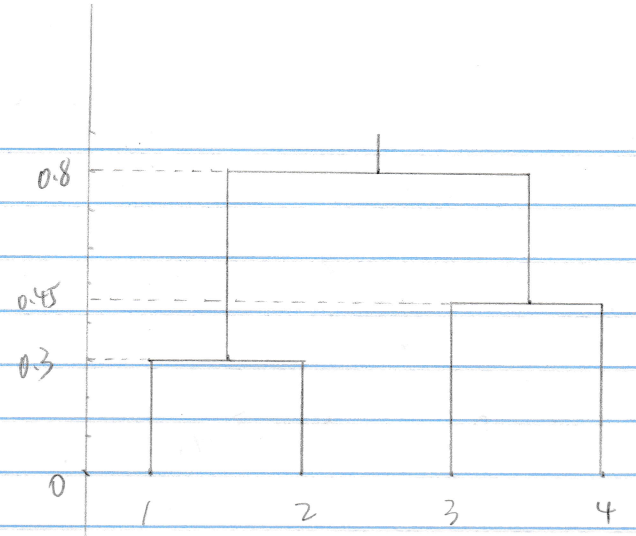
Thus the identity is proved.

(b) Based on the identity, we see that minimizing the within-cluster variance (right-hand side) is equivalent to minimizing the sum of all pairwise squared Euclidean distance (left-hand side, and the objective).

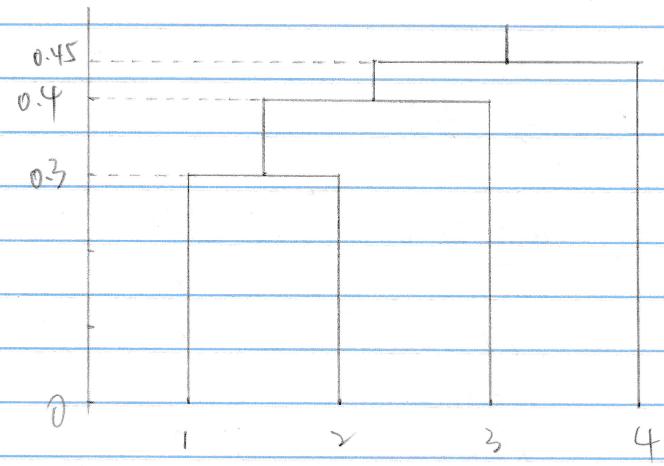
In step 2-b of algorithm 10.1, when we re-assign each observation to the cluster whose centroid is closest, we are trying to minimize within-cluster variance. Thus it decreases the objective at each iteration.

2.

(a)



(b)



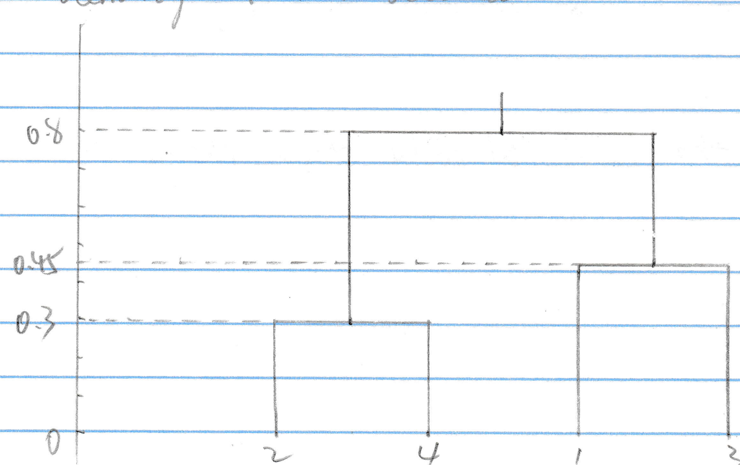
(c) Cluster 1: 1, 2 Cluster 2: 3, 4

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

(e) Lets re-write the dissimilarity matrix as:

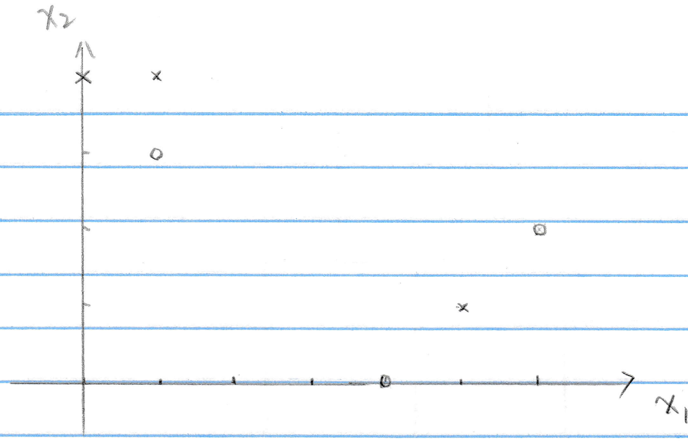
$$\begin{array}{cc} & \begin{array}{cccc} 4 & 1 & 3 & 2 \end{array} \\ \begin{array}{c} 4 \\ 1 \\ 3 \\ 2 \end{array} & \begin{bmatrix} & 0.7 & 0.45 & 0.8 \\ 0.7 & & 0.4 & 0.3 \\ 0.45 & 0.4 & & 0.5 \\ 0.8 & 0.3 & 0.5 & \end{bmatrix} \end{array}$$

The dendrogram then becomes:



3.

(a)



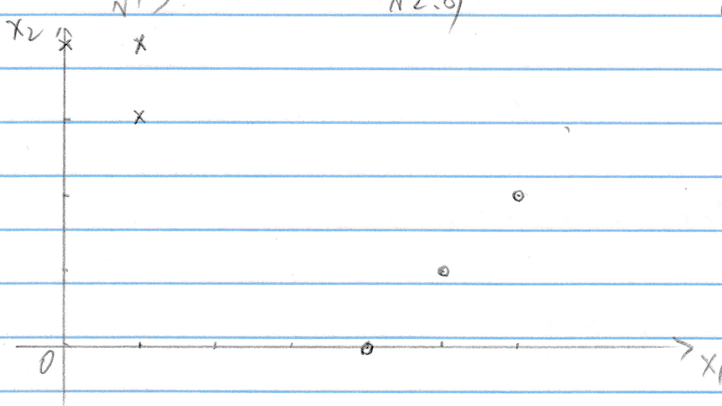
(b) "x" and "o" are assigned.

(c) "x" cluster's centroid: $(2, 3)$

"o" cluster's centroid: $(\frac{11}{3}, \frac{5}{3})$

(d)

obs	dist to "x"	dist to "o"	label
1.	$\sqrt{2}$	$\sqrt{12.56}$	x
2	$\sqrt{1}$	$\sqrt{8.89}$	x
3	$\sqrt{5}$	$\sqrt{18.89}$	x
4	$\sqrt{13}$	$\sqrt{2.22}$	o
5	$\sqrt{17}$	$\sqrt{5.55}$	o
6.	$\sqrt{13}$	$\sqrt{2.89}$	o.



(e) After 1 iteration, the label already stops changing.

(d) (omitted)

4.

1(a) The complete linkage dendrogram will occur higher. X

Consider : all pairwise dissimilarity (distance) are equal.

If they are not equal, single likage will generally be lower.

(b) They'll fuse at the same height, since for single-clusters,
single linkage = complete linkage

5. Left: Group retailers with similar sock data together.
Center: Group retailers with both similar sock data and similar computer data together.
Right: Group retailers with similar computer data together.

6.

(a) For each observation, it can be projected to the first component vector and gets a score. The variance of that score takes 10% of the total variance when we consider all principal components.

(b) Add the machine used (A vs. B) as a feature of the dataset.

(c) (See code).