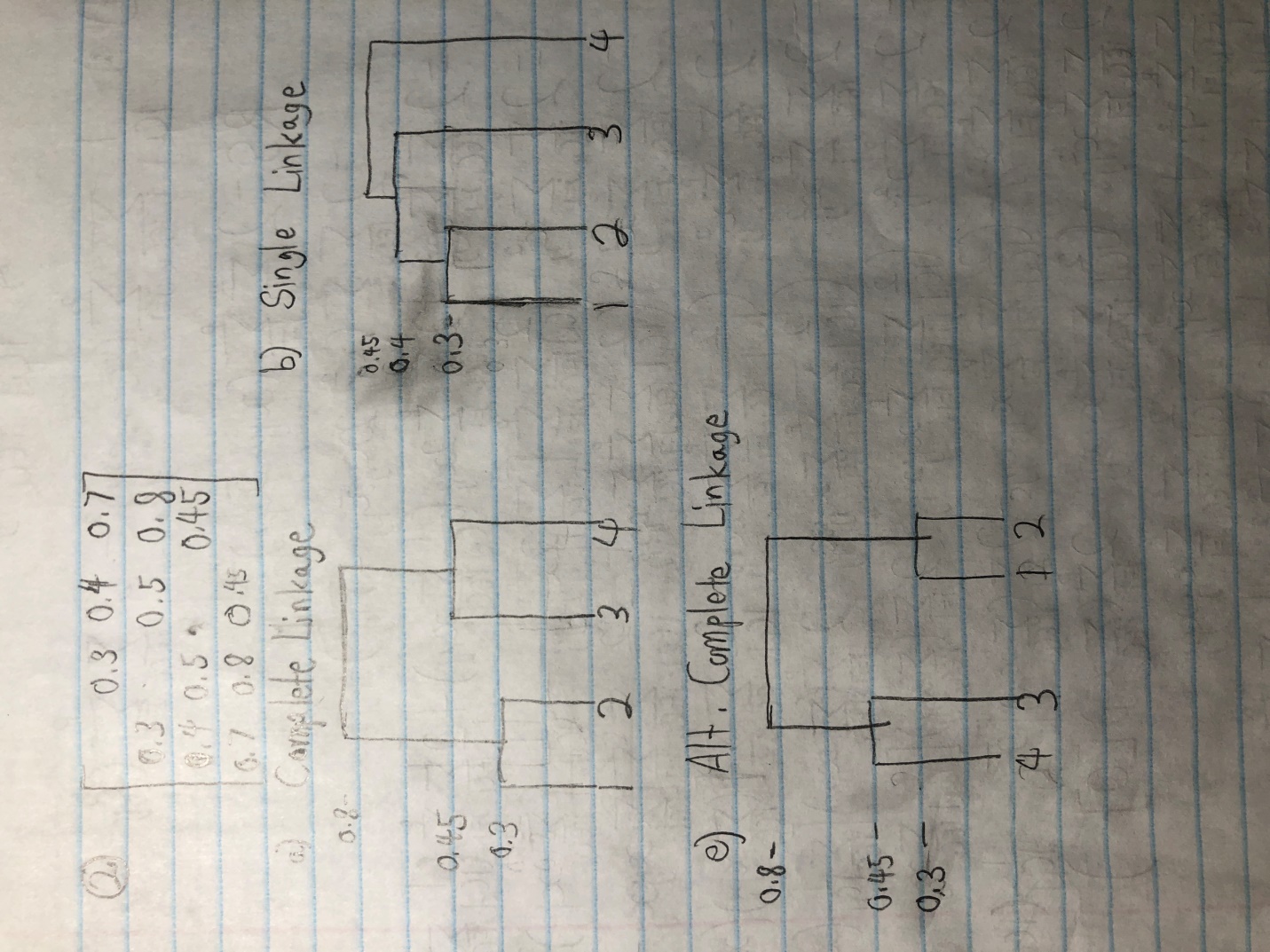
1. K-Clustering Algorithm
   1. Proof that . Note that .

Therefore, we have shown that , as desired.

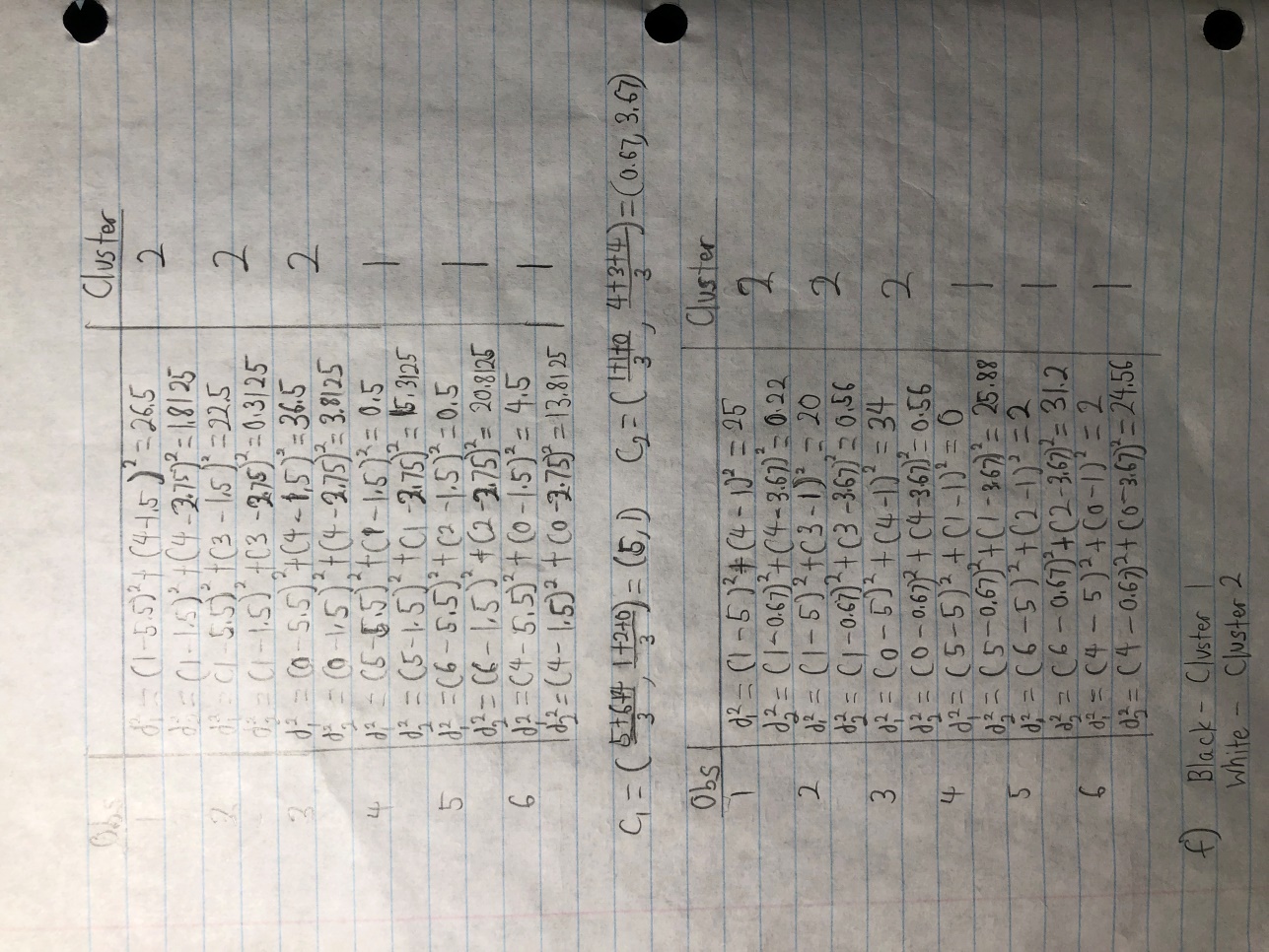
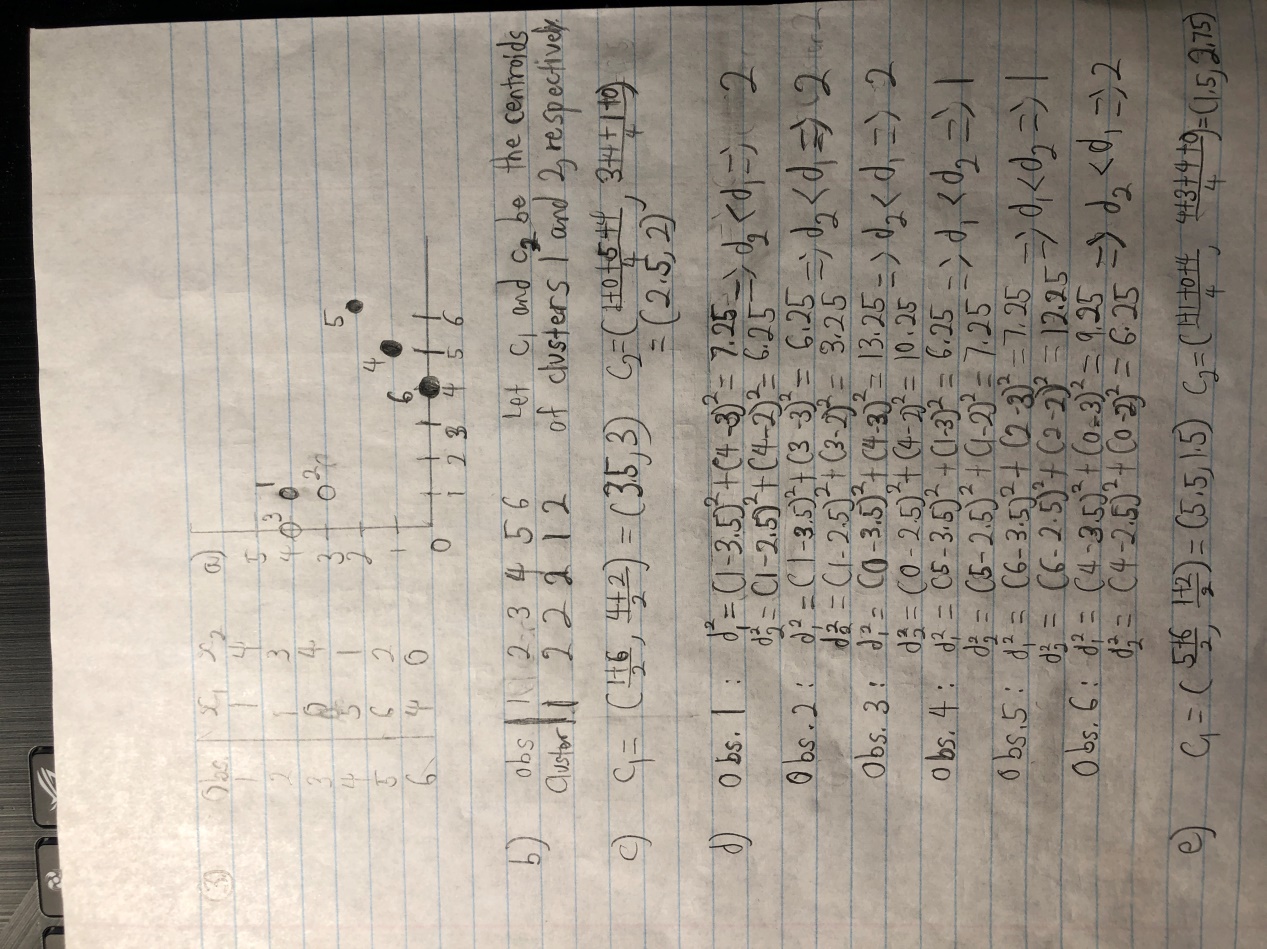
* 1. As we know, Algorithm 10.1 first randomly assigns each observation to a cluster, then repeatedly computes each of the cluster centroids (defined as the vector of the *p* predictor means for the observations in that cluster),assigning every observation to the cluster whose centroid has the least Euclidean distance between it and the observation on each iteration. As we can see in identity 10.12, the sum of all the pairwise squared Euclidean distances in a given cluster as defined in identity 10.10 is equivalent to 2 times the sum of the squared Euclidean distances between each observation and that given cluster’s centroid (as defined earlier). Therefore, the sum of the pairwise squared Euclidean distances in a given cluster is directly proportional to 2 times the sum of the squared Euclidean distances between each observation of the cluster and the cluster’s mean. Algorithm 10.1’s 2b step attempts to minimize the distances (distance defined as squared Euclidean distance) between each observation and its cluster centroid, which in turn minimizes the pairwise distances in each cluster, since these two sums are proportional to each other. This thereby decreases the objective 10.11 as well, as 10.11 is also attempting to minimize the pairwise distances in each cluster. On each iteration after the first, the 2a step of the algorithm updates the cluster centroids accordingly to match with the newly formed clusters from the 2b step in the previous iteration, before proceeding to perform the 2b step once more in the current iteration. Therefore, it can be concluded that the K-means clustering algorithm 10.1 decreases the objective 10.11 at each of its iterations.

1. Dendograms



1. If the dendogram from a) is cut into two clusters, then {1,2} and {3,4} would be the result.
2. If the dendogram from b) is cut into two clusters, then {1,2,3} and {4} would be the result.
3. Manual K-Clustering

a), b), c), d), first part of e)

e), continued

1. Hierarchical Clustering Analysis
   1. There isn’t enough information to tell if the clusters {1,2,3} and {4,5} will fuse at the same point in both a single linkage dendrogram and a complete linkage dendrogram or not. While it is possible that the maximum and minimum dissimilarities between the two clusters are different (which would lead to them fusing at different heights on the two dendrograms), it is also possible that the maximum and minimum dissimilarities between the two are the same, which occurs if and only if every dissimilarity between every pair of observations, formed by taking one from each cluster, was equal.
   2. The clusters {5} and {6} will fuse at the same point in both a complete linkage dendrogram and a single linkage dendrogram. This is because of the fact that both clusters consist of a single observation, meaning that the maximum and minimum dissimilarities between the two clusters must be the same (there is only ever one dissimilarity in this case).
2. Examining Figure 10.14
   1. For the first scaling, K-means clustering would likely cluster the 4 shoppers with the lowest number of purchased socks into one cluster, with the other 4 being in the other cluster. In terms of Euclidean distance, the fact that the computer purchases have such small values as predictors compared to sock purchases means that they will likely not impact K-means clustering much or at all, due to how K-means clustering relies on the Euclidean distance to determine which clusters are formed.
   2. For the second scaling, K-means clustering would probably weigh the computer purchases much more than in the first scaling, making it likely that it simply makes 2 clusters based on who purchased a computer and who did not.
   3. For the third scaling, K-means clustering would almost certainly place the 4 shoppers who purchased a computer into one cluster, and place the 4 shoppers who did not purchase a computer into the other. This version of the data has been scaled such that the sock purchases have almost no weight when compared to the computer purchases, so the Euclidean distances between the observations depend entirely on the latter.
3. Gene Expression Measurements
   1. The fact that the first principal component “explains 10% of the variation” implies that 10% of information about the relationships between the variables is contained in the first principal component. It also implies that 90% of the information in the dataset is omitted in the projection onto the first principal component.
   2. The main issue with this approach is that the first principal component barely explains any of the information in the dataset, and that there is clearly a time-based trend. This is evident because there is a strong-linear left-right trend, and it is stated that the researcher used one machine more in earlier times and the other more for later times. Since he has a record of which samples were run on which machines, it would be advisable to add in those records as a predictor in the dataset, as including the strong time-based trend would likely improve the amount of variability explained by the first principal component.
   3. In R file