Name: Jingyi Yuan
UNI: jy2736
Class: STAT W4240

Homework 06

Question 1

We choose the exponential function: $L(y, f(x)) = \exp(-y f(x))$

So at iteration m-1, our boosted classifier is the solution of:

$$(\beta_m, G_m) = \underset{\beta, G}{\arg\min} \sum_{i=1}^{N} L[y_i, f_{(m-1)}(x_i) + \beta G(x_i)]$$

$$= \underset{\beta, G}{\arg\min} \sum_{i=1}^{N} \exp[-y_i \cdot (f_{m-1}(x_i) + \beta G(x_i))]$$

This can be expressed as $(\beta_m, G_m) = \underset{\beta, G}{\arg\min} \sum_{i=1}^{N} w_i^{(m)} \exp(-\beta y_i G(x_i))$

with $w_i^{(m)} = \exp(-y_i f_{m-1}(x_i))$

The solution can be obtained in two steps. First, for any value of $\beta > 0$, the solution for $G_m(x)$ is

$$G_m = \underset{G}{\arg\min} \sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G(x_i))$$

which is the classifier that minimizes the weighted error rate in predicting y

This can be easily seen by expressing the criterion as:

$$e^{-\beta} \sum_{y_i = G(x_i)} w_i^{(m)} + e^{\beta} \sum_{y_i \neq G(x_i)} w_i^{(m)}$$

$$\Rightarrow (e^{\beta} - e^{-\beta}) \sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G(x_i)) + e^{-\beta} \sum_{i=1}^{N} w_i^{(m)}$$

We take the derivative with respect to β:

$$(e^{\beta} + e^{-\beta}) \cdot \sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G(x_i)) - e^{-\beta} \sum_{i=1}^{N} w_i^{(m)} = 0.$$

$$(e^{2\beta} + 1) \sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G(x_i)) - \sum_{i=1}^{N} w_i^{(m)} = 0$$

$$\therefore e^{2\beta} = \frac{\sum_{i=1}^{N} w_i^{(m)}}{\sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G(x_i))} - 1$$

Let's suppose that $err_m = \frac{\sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G(x_i))}{\sum_{i=1}^{N} w_i^{(m)}} - 1$

$$\therefore e^{2\beta} = \frac{1}{err_m} - 1 = \frac{1 - err_m}{err_m}$$

$$\therefore \beta = \frac{1}{2} \ln\left(\frac{1 - err_m}{err_m}\right)$$

We have derived the expression (10.12) for the update parameter in Adaboost.

Question 2

$$f^*(x) = \underset{f(x)}{\arg\min}\ E_{Y|x}(e^{-Yf(x)}).$$

$$\frac{\partial E_{Y|x}(e^{-Yf(x)})}{\partial f(x)} = E_{Y|x}(-Y \cdot e^{-Yf(x)})$$

$$= -e^{-f(x)} Pr(Y=1|x) + e^{f(x)} \cdot Pr(Y=-1|x) = 0$$

$$\Rightarrow \ -Pr(Y=1|x) + e^{2f(x)} \cdot Pr(Y=-1|x) = 0$$

$$\therefore e^{2f(x)} = \frac{Pr(Y=1|x)}{Pr(Y=-1|x)}$$

$$\therefore f^*(x) = f(x) = \frac{1}{2} \log \frac{Pr(Y=1|x)}{Pr(Y=-1|x)}.$$

We have proved result (10.6).

Question 3

$$\hat{\gamma}_{jm} = \underset{\gamma_{jm}}{\arg\min} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma_{jm})$$

$$= \underset{\gamma_{jm}}{\arg\min} \sum_{x_i \in R_{jm}} \exp[-y_i(f_{m-1}(x_i) + \gamma_{jm})]$$

This can be expressed as $\hat{\gamma}_{jm} = \underset{\gamma_{jm}}{\arg\min} \sum_{x_i \in R_{jm}} w_i^{(m)} \cdot \exp(-y_i \gamma_{jm})$

with $w_i^{(m)} = \exp(-y_i f_{m-1}(x_i))$.

We take the derivative with respect to γ_{jm}

$$\sum_{x_i \in R_{jm}} (-y_i) \cdot w_i^{(m)} \exp(-y_i \gamma_{jm}) = 0$$

$$\therefore -\sum_{x_i \in R_{jm}} w_i^{(m)} \cdot \exp(-\gamma_{jm}) I(y_i = 1) + \sum_{x_i \in R_{jm}} w_i^{(m)} \exp(\gamma_{jm}) I(y_i = -1) = 0.$$

$$\therefore e^{2\gamma_{jm}} \sum_{x_i \in R_{jm}} w_i^{(m)} I(y_i = -1) - \sum_{x_i \in R_{jm}} w_i^{(m)} I(y_i = 1) = 0$$

$$\therefore e^{2\gamma_{jm}} = \frac{\sum_{x_i \in R_{jm}} w_i^{(m)} I(y_i = 1)}{\sum_{x_i \in R_{jm}} w_i^{(m)} I(y_i = -1)}$$

$$\therefore \hat{\gamma}_{jm} = \frac{1}{2} \log \frac{\sum_{x_i \in R_{jm}} w_i^{(m)} I(y_i = 1)}{\sum_{x_i \in R_{jm}} w_i^{(m)} I(y_i = -1)}.$$

Question 4

(a).

$$\text{left} = \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2$$

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

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

$$= 2 \sum_{i \in C_k} \sum_{j=1}^{p} x_{ij}^2 - 2 \sum_{i \in C_k} \sum_{j=1}^{p} \bar{x}_{kj} \cdot x_{ij}$$

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

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

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

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

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

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

We have proved (10.12)

(b).

In Algorithm 10.1 Step 2(a) the cluster means for each feature are the constants that minimize the sum-of-squared deviations. In Step 2(b), we assign each observation to the cluster whose centroid is closet, and reallocating the observations can only improve the right formula and its value decreases, which also decreases the left one. This means that as the algorithm is run, the clustering obtained will continually improve until the result no longer changes; the objective of the left formula will never increase. When the result no longer changes, a local optimum has been reached.

Question 5

(a).

$$\begin{pmatrix} & 0.3 & 0.4 & 0.7 \\ 0.3 & & 0.5 & 0.8 \\ 0.4 & 0.5 & & 0.45 \\ 0.7 & 0.8 & 0.45 & \end{pmatrix}$$

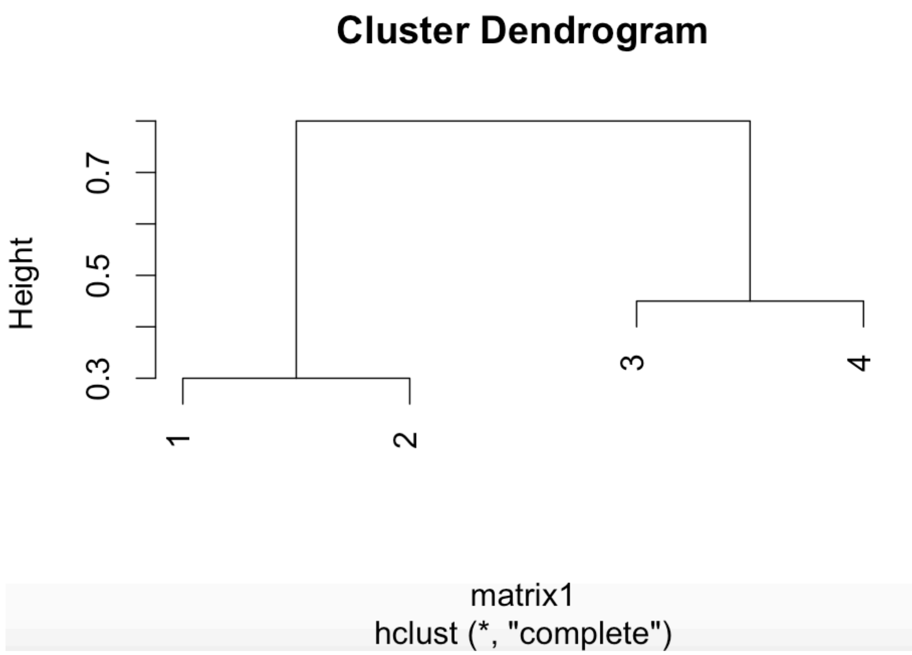
The minimum distance of this dissimilarity matrix is 0.3, which is between the first and second observation, so we form a group {1,2}. Then we have:

$$\begin{pmatrix} & 0.5 & 0.8 \\ 0.5 & & 0.45 \\ 0.8 & 0.45 & \end{pmatrix}$$

The minimum distance of this dissimilarity matrix is 0.45, which is between the third and the forth observation, so we form a group {3,4}. Then we have:

$$\begin{pmatrix} & 0.8 \\ 0.8 & \end{pmatrix}$$

Finally we form a group {{1,2},{3,4}} and the distance is 0.8.



(b).

$$\begin{pmatrix} & 0.3 & 0.4 & 0.7 \\ 0.3 & & 0.5 & 0.8 \\ 0.4 & 0.5 & & 0.45 \\ 0.7 & 0.8 & 0.45 & \end{pmatrix}$$

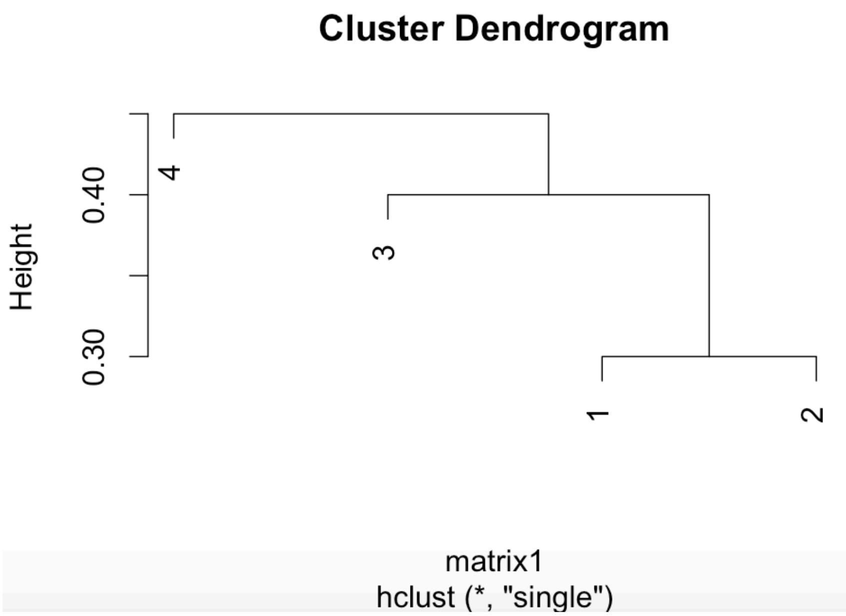
The minimum distance of this dissimilarity matrix is 0.3, which is between the first and second observation, so we form a group {1,2}. Then we have:

$$\begin{pmatrix} & 0.4 & 0.7 \\ 0.4 & & 0.45 \\ 0.7 & 0.45 & \end{pmatrix}$$

The minimum distance of this dissimilarity matrix is 0.4, which is between the group {1,2} and the third observation, so we form a group {{1,2},3}. Then we have:

$$\begin{pmatrix} & 0.45 \\ 0.45 & \end{pmatrix}$$

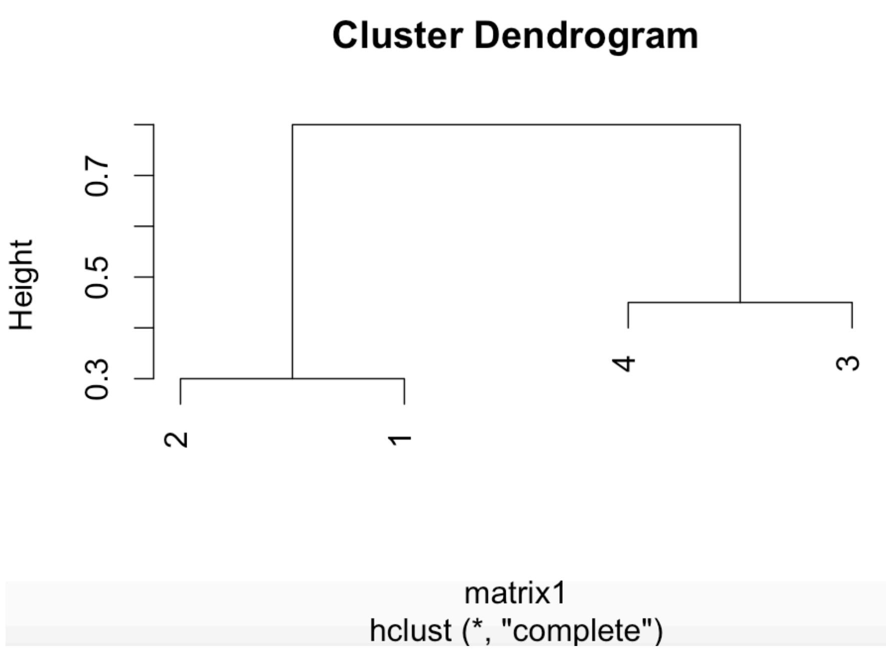
Finally we form a group {{{1,2},3},4} and the distance is 0.45.



(c).
We have clusters {1,2} and {3,4}.

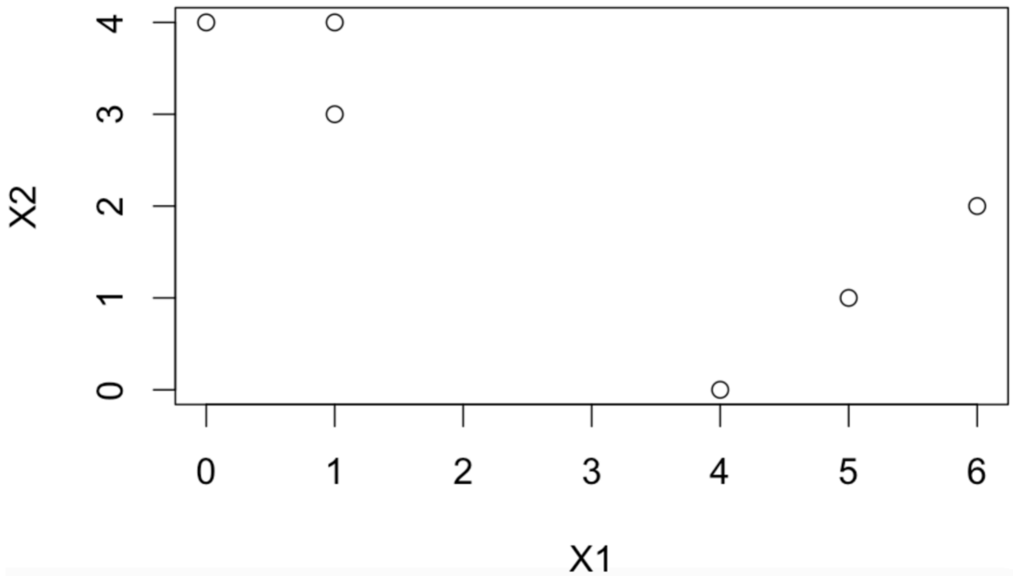
(d).
We have clusters {{1,2},3} and {4}.

(e).



Question 6

(a).

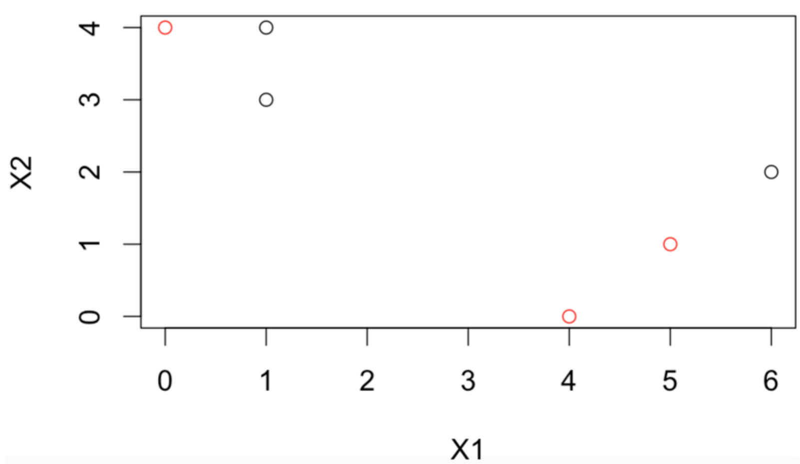


(b).

The cluster labels for each observation is shown below:

```
> labels
[1] 1 1 2 2 1 2
```

“1” for black and “2” for red thus we have a labeled picture:



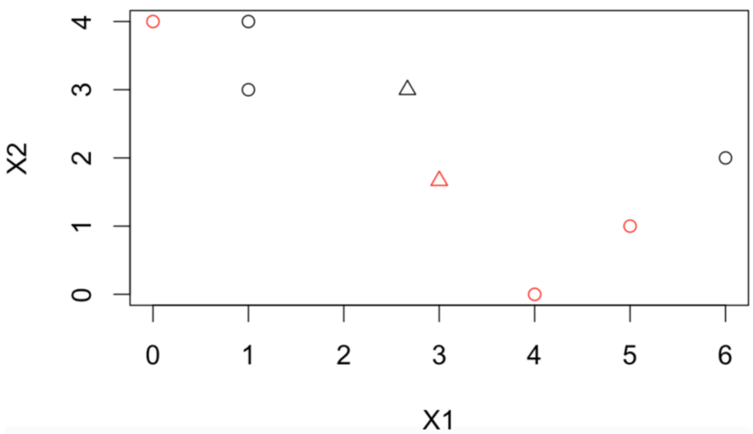
(c).

We have centroid1 and centroid2 of label “1” and ”2” shown as below:

```
> centroid1
[1] 2.666667 3.000000
> centroid2
[1] 3.000000 1.666667
```

(d).

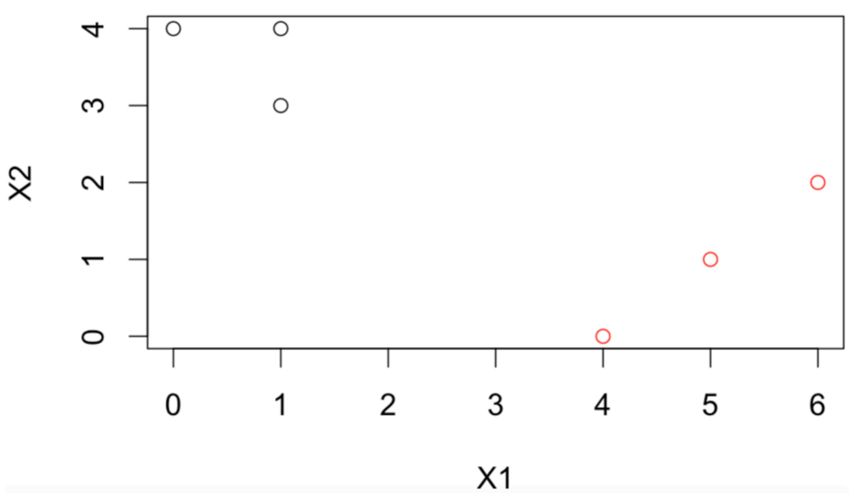
We draw the centroids on the picture and use the Euclidean distance to find the closest centroid to each observation and have another vector of labels:



(the centroids are in triangle)

```
labels_d = c(1, 1, 1, 2, 2, 2)
```

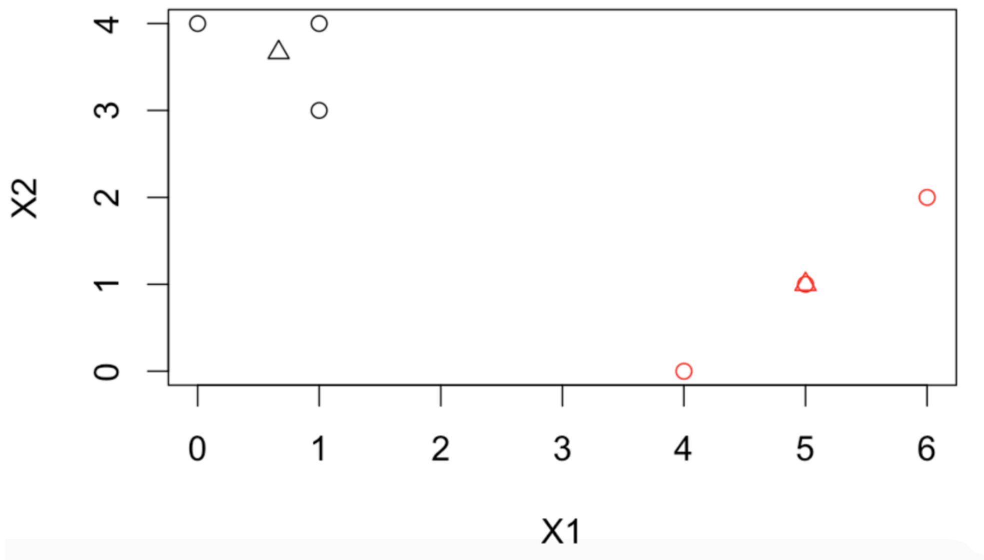
We re-assign the color of each point according to the new labels:



(e).

We repeat the former steps and find the new centroids, plot them on the pictures and re-colored each point until the answers obtained stop changing.

```
> centroid1
[1] 0.6666667 3.6666667
> centroid2
[1] 5 1
```



(f).