SOLUTIONS TO PRACTICE PROBLEMS

1. (a) The estimated correlation matrix from the one factor model is

$$\boldsymbol{\ell}\boldsymbol{\ell}^T + \Psi = \begin{pmatrix} 1.00 & 0.49 & 0.49 & 0.60 & 0.60 \\ 0.49 & 1.00 & 0.61 & 0.75 & 0.75 \\ 0.49 & 0.61 & 1.00 & 0.75 & 0.75 \\ 0.60 & 0.75 & 0.75 & 1.00 & 0.91 \\ 0.60 & 0.75 & 0.75 & 0.91 & 1.00 \end{pmatrix}$$

where the diagonal elements of Ψ are the uniquenesses $0.605, 0.383, 0.389, 0.084, 0.089$ while

$$\boldsymbol{\ell} = \begin{pmatrix} 0.628 \\ 0.786 \\ 0.781 \\ 0.957 \\ 0.954 \end{pmatrix}$$

The p-value for the test of the hypothesis that 1 factor is sufficient is 0.761, which suggests that the 1 factor model is OK.

(b) The matrix Q is orthogonal ($QQ^T = I$) and the new loadings are

$$\boldsymbol{\ell}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0.335 - 0.939 \\ 0.734 - 0.280 \\ 0.757 - 0.206 \\ 0.915 - 0.299 \\ 0.886 - 0.343 \end{pmatrix}$$

and

$$\boldsymbol{\ell}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0.335 + 0.939 \\ 0.734 + 0.280 \\ 0.757 + 0.206 \\ 0.915 + 0.299 \\ 0.886 + 0.343 \end{pmatrix}$$

2. (a) By symmetry, $P(X_1 > 0) = 1/2$ for both densities. Therefore,

$$P(\text{misclassification}) = \frac{1}{3}P(X_1 > 0|f_1) + \frac{2}{3}P(X_1 < 0|f_2) = \frac{1}{2}$$

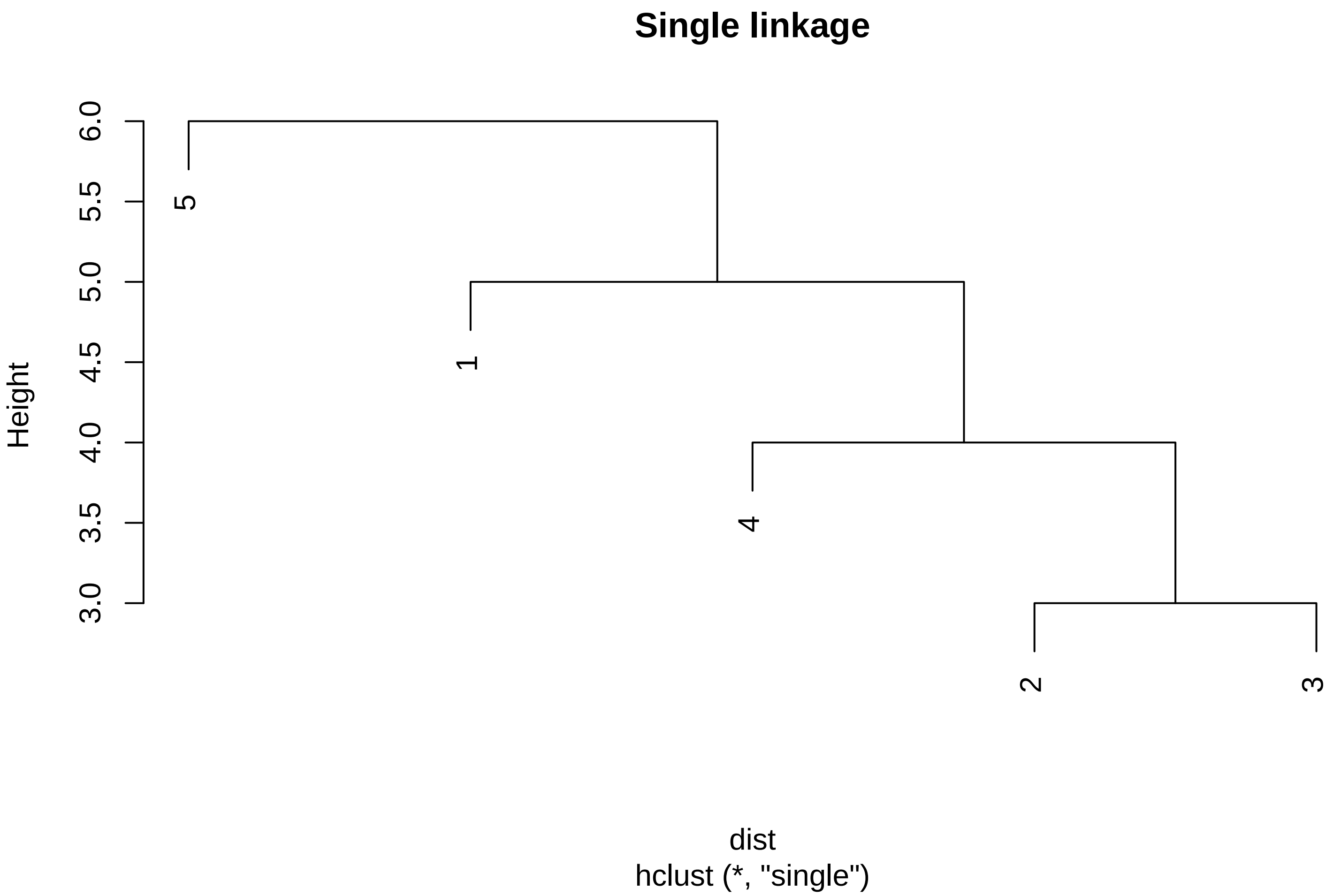
(b) The optimal classification rule classifies an observation $\boldsymbol{x}$ as belonging to f_1 if $f_1(\boldsymbol{x})/3 > 2f_2(\boldsymbol{x})/3$ and as belonging to f_2 if $2f_2(\boldsymbol{x})/3 > f_1(\boldsymbol{x})/3$. In other words, the optimal rule depends on whether $(x_1^2 + x_2^3$ is greater than or less than $1/4$.

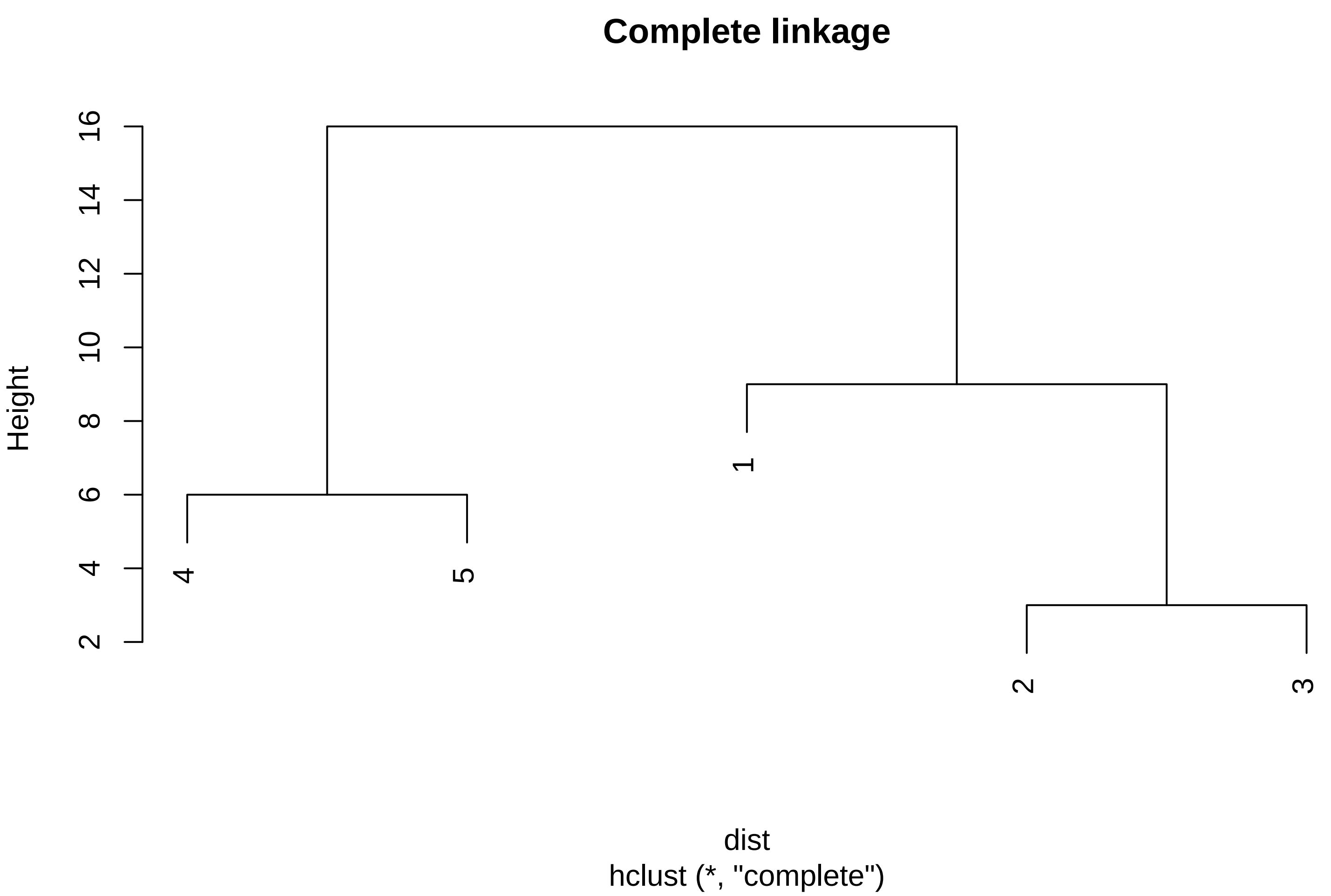
(c) Knowing the signs of X_1 and X_2 tells us only which quadrant the observation lies in. If we define $Y_1 = \text{sign}(X_1)$ and $Y_2 = \text{sign}(X_2)$ then the distribution of (Y_1, Y_2) is the same under f_1 and f_2. The optimal classification rule in this case is to always classify as belonging to f_2, with a misclassification rate of 1/3.

3. PCA effectively does an eigen-decomposition of the correlation or covariance matrix of the data to define new uncorrelated variables, the principal components. In many case, the first few PCs (those with largest variance) will describe the structure of the data very well. The model for ICA is $\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Y}$ where the components of $\boldsymbol{Y}$ are assumed to be independent, not simply uncorrelated. Like PCA, ICA can be used for dimension reduction but more generally, we are interested in all the components of the random vector $\boldsymbol{Y}$.

4. In the factor analysis model, the correlation matrix $R = LL^T + \Psi$. If $\widehat{R}$ is the sample correlation matrix then $\widehat{R} = V\Lambda V^T$ where the diagonal elements of Λ are ordered from largest to smallest. Therefore, if we want an r factor model, we can define L to be the first r columns of V multiplied by the square roots of the eigenvalues $\lambda_1^{1/2} \geq \cdots \geq \lambda_r^{1/2}$ and Ψ to be a diagonal matrix so that the diagonal elements of $LL^T + \Psi$ are 1s.
(The same can be done with the sample covariance matrix making appropriate modfications to the argument.)

5. (a) The first cluster will contain observations 2 and 3 (whose distance is 3).
(b) See plots attached.


Single linkage
Height
6.0
5.5
5.0
4.5
4.0
3.5
3.0
5
1
4
2
3
dist
hclust (*, "single")


Complete linkage
Height
16
14
12
10
8
6
4
2
4
5
1
2
3
dist
hclust (*, "complete")