Prof. Dr. Rudolf Mathar, Dr. Arash Behboodi, Markus Rothe

Exercise 10
- Proposed Solution -

Friday, January 11, 2019

Solution of Problem 1

Each calculated distance $d(\mathbf{x}, \mathbf{y})$ is equivalent to 0.5P. The updated centers in **b)** each is equivalent to 0.5P.

a) The center of cluster 1 is $\mathbf{c}_1 = \mathbf{x}_1$, and the center of cluster 2 is $\mathbf{c}_2 = \mathbf{x}_3$.

$$d(\mathbf{c}_1, \mathbf{x}_2) = \sqrt{(7-7)^2 + (3-0)^2} = 3$$
$$d(\mathbf{c}_2, \mathbf{x}_2) = \sqrt{(9-7)^2 + (1-3)^2} = \sqrt{8} = 2.8284$$

$\mathbf{x}_2$ belongs to cluster 2.

$$d(\mathbf{c}_1, \mathbf{x}_4) = \sqrt{(9-7)^2 + (5-0)^2} = \sqrt{29} = 5.38$$
$$d(\mathbf{c}_2, \mathbf{x}_4) = \sqrt{(9-9)^2 + (5-1)^2} = 4.$$

$\mathbf{x}_4$ belongs to cluster 2.

$$d(\mathbf{c}_1, \mathbf{x}_5) = \sqrt{(3-7)^2 + (7-0)^2} = \sqrt{65} = 8.06$$
$$d(\mathbf{c}_2, \mathbf{x}_5) = \sqrt{(9-3)^2 + (1-7)^2} = \sqrt{72} = 8.485.$$

$\mathbf{x}_5$ belongs to cluster 1.

$$d(\mathbf{c}_1, \mathbf{x}_6) = \sqrt{(12-7)^2 + (3-0)^2} = \sqrt{34} = 5.83$$
$$d(\mathbf{c}_2, \mathbf{x}_6) = \sqrt{(12-9)^2 + (3-1)^2} = \sqrt{13} = 3.605.$$

$\mathbf{x}_6$ belongs to cluster 2.

b) The new center of cluster 1 is

$$\left(\tfrac{7+3}{2}, \tfrac{0+7}{2}\right) = (5, 3.5)$$

The new center of cluster 2 is

$$\left(\tfrac{7+9+9+12}{4}, \tfrac{3+1+5+3}{4}\right) = (9.25, 3).$$

c) Using $d_1(\mathbf{x}, \mathbf{y})$:

$$d(\mathbf{c}_1, \mathbf{x}_2) = |7 - 7| + |3 - 0| = 3$$
$$d(\mathbf{c}_2, \mathbf{x}_2) = |9 - 7| + |1 - 3| = 4$$

$\mathbf{x}_2$ belongs to cluster 1.

$$d(\mathbf{c}_1, \mathbf{x}_4) = |9 - 7| + |5 - 0| = 7$$
$$d(\mathbf{c}_2, \mathbf{x}_4) = |9 - 9| + |5 - 1| = 4.$$

$\mathbf{x}_4$ belongs to cluster 2.

$$d(\mathbf{c}_1, \mathbf{x}_5) = |3 - 7| + |7 - 0| = 11$$
$$d(\mathbf{c}_2, \mathbf{x}_5) = |9 - 3| + |1 - 7| = 12.$$

$\mathbf{x}_5$ belongs to cluster 1.

$$d(\mathbf{c}_1, \mathbf{x}_6) = |12 - 7| + |3 - 0| = 8$$
$$d(\mathbf{c}_2, \mathbf{x}_6) = |12 - 9| + |3 - 1| = 5.$$

$\mathbf{x}_6$ belongs to cluster 2.

Using $d_\infty(\mathbf{x}, \mathbf{y})$:

$$d(\mathbf{c}_1, \mathbf{x}_2) = \max(\mid 7 - 7 \mid, \mid 3 - 0 \mid) = \max(0, 3) = 3$$
$$d(\mathbf{c}_2, \mathbf{x}_2) = \max(\mid 9 - 7 \mid, \mid 1 - 3 \mid) = \max(2, 2) = 2$$

$\mathbf{x}_2$ belongs to cluster 2.

$$d(\mathbf{c}_1, \mathbf{x}_4) = \max(\mid 9 - 7 \mid, \mid 5 - 0 \mid) = \max(2, 5) = 5$$
$$d(\mathbf{c}_2, \mathbf{x}_4) = \max(\mid 9 - 9 \mid, \mid 5 - 1 \mid) = \max(0, 4) = 4.$$

$\mathbf{x}_4$ belongs to cluster 2.

$$d(\mathbf{c}_1, \mathbf{x}_5) = \max(\mid 3 - 7 \mid, \mid 7 - 0 \mid) = \max(4, 7) = 7$$
$$d(\mathbf{c}_2, \mathbf{x}_5) = \max(\mid 3 - 9 \mid, \mid 7 - 1 \mid) = \max(6, 6) = 6.$$

$\mathbf{x}_5$ belongs to cluster 2.

$$d(\mathbf{c}_1, \mathbf{x}_6) = \max(\mid 12 - 7 \mid, \mid 3 - 0 \mid) = \max(5, 3) = 5$$
$$d(\mathbf{c}_2, \mathbf{x}_6) = \max(\mid 12 - 9 \mid, \mid 3 - 1 \mid) = \max(3, 2) = 3.$$

$\mathbf{x}_6$ belongs to cluster 2.

Solution of Problem 2

a) If the n points are clustered into $S_1, \ldots, S_g$, then ML-cluster analysis writes as

$$\max_{S_1,\ldots,S_g} \sum_{k=1}^{g} \sum_{i\in S_k} \log f_k(\mathbf{x}_i) = \max_{S_1,\ldots,S_g} \sum_{k=1}^{g} \sum_{i\in S_k} \mathsf{const.} - \frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}\left\{(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \mathbf{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\right\}.$$

Therefore having Σ and μ_k, the ML-cluster analysis is given by

$$\min_{S_1,\ldots,S_g} \sum_{k=1}^{g} \sum_{i\in S_k} \log|\mathbf{\Sigma}| + \left\{(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \mathbf{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)\right\}.$$

b) Given clustering of samples $S_1, \ldots, S_g$, the ML-estimation of $\mathbf{\Sigma}$ results from the minimization of above expression for fixed $S_1, \ldots, S_g$. Following similar argument from ML estimation of covariance matrix, μ_k are estimated by $\overline{\mathbf{x}}_k$. Using these values and differentiating with respect to $\mathbf{\Sigma}^{-1}$, similar to ML-estimation of covariance matrix, the ML-estimation of $\mathbf{\Sigma}$ is given by:

$$n\hat{\mathbf{\Sigma}} = \sum_{k=1}^{g} \sum_{i\in S_k} \left\{(\mathbf{x}_i - \overline{\mathbf{x}}_k)(\mathbf{x}_i - \overline{\mathbf{x}}_k)^T\right\} \implies \hat{\mathbf{\Sigma}} = \frac{1}{n}\mathbf{W},$$

where $\mathbf{W}$ is within-group sum of squares.

c) Using the above estimation, ML-estimation can be written as

$$\min_{S_1,\ldots,S_g} \sum_{k=1}^{g} \sum_{i\in S_k} \log|\frac{\mathbf{W}}{n}| + \left\{(\mathbf{x}_i - \overline{\mathbf{x}}_k)^T \mathbf{W}^{-1} n(\mathbf{x}_i - \overline{\mathbf{x}}_k)\right\}.$$

But:

$$\sum_{k=1}^{g} \sum_{i\in S_k} (\mathbf{x}_i - \overline{\mathbf{x}}_k)^T \mathbf{W}^{-1} n(\mathbf{x}_i - \overline{\mathbf{x}}_k) = \sum_{k=1}^{g} \sum_{i\in S_k} \mathrm{tr}(n\mathbf{W}^{-1}(\mathbf{x}_i - \overline{\mathbf{x}}_k)(\mathbf{x}_i - \overline{\mathbf{x}}_k)^T) = n^2 \mathrm{tr}(\mathbf{W}^{-1}\mathbf{W}) = n^2 \cdot p.$$

In other words, the second term in the optimization problem is constant and thus not relevant for the optimization problem. We remove it from the optimization problem. Also, the division by n in the first term does not change the optimization problem and logarithm can be left out from the optimization problem, too. Therefore, the ML-estimation can be written as:

$$\min_{S_1,\ldots,S_g} \det(\mathbf{W}).$$

d) If $\mathbf{\Sigma}$ is known, ML-cluster analysis is written as:

$$\min_{S_1,\ldots,S_g} \sum_{k=1}^{g} \sum_{i\in S_k} \log|\mathbf{\Sigma}| + \left\{(\mathbf{x}_i - \overline{\mathbf{x}}_k)^T \mathbf{\Sigma}^{-1}(\mathbf{x}_i - \overline{\mathbf{x}}_k)\right\}.$$

Since Σ is known and irrelevant for the optimization, only the second term is important. Now see that from the argument used above:

$$\sum_{k=1}^{g} \sum_{i\in S_k} (\mathbf{x}_i - \overline{\mathbf{x}}_k)^T \mathbf{\Sigma}^{-1}(\mathbf{x}_i - \overline{\mathbf{x}}_k) = \mathrm{tr}(\mathbf{W}\mathbf{\Sigma}^{-1}).$$

Therefore the ML-analysis writes as:

$$\min_{S_1,\ldots,S_g} \mathrm{tr}(\mathbf{W}\mathbf{\Sigma}^{-1}).$$

Solution of Problem 3

a) Since the number of data points in each class is the same we have that

$$\mathbf{E}_1 = \mathbf{E}_2 = \mathbf{I}_3 - \frac{1}{3}\mathbf{1}_3\mathbf{1}_3^{\mathrm{T}} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

b)

$$\bar{\mathbf{x}} = \frac{1}{6}\sum_{k=1}^{6}\mathbf{x}_k = \frac{1}{6}\begin{bmatrix} -1+1+2+1+0-1 \\ 1-2+0+1+2+1 \\ 1+0-1-1-1-1 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 2 \\ 3 \\ -3 \end{bmatrix}$$

c)

$$\bar{\mathbf{x}}_1 = \frac{\mathbf{x}_1+\mathbf{x}_2+\mathbf{x}_3}{3} = \frac{1}{3}\begin{bmatrix} -1+1+2 \\ 1-2+0 \\ 1+0-1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 4 \\ -2 \\ 0 \end{bmatrix},$$

$$\bar{\mathbf{x}}_2 = \frac{\mathbf{x}_4+\mathbf{x}_5+\mathbf{x}_6}{3} = \frac{1}{3}\begin{bmatrix} 1+0-1 \\ 1+2+1 \\ -1-1-1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 0 \\ 4 \\ -3 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 0 \\ 8 \\ -6 \end{bmatrix}$$

d)

$$\bar{\mathbf{x}}_1 - \bar{\mathbf{x}} = \frac{1}{6}\begin{bmatrix} 2 \\ -5 \\ 3 \end{bmatrix}, \qquad \bar{\mathbf{x}}_2 - \bar{\mathbf{x}} = \frac{1}{6}\begin{bmatrix} -2 \\ 5 \\ -3 \end{bmatrix}$$

Note that $(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}) = -(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}})$, therefore we have that

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}})(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}})^{\mathrm{T}} = (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}})(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}})^{\mathrm{T}} = \frac{1}{6^2}\begin{bmatrix} 4 & -10 & 6 \\ -10 & 25 & -15 \\ 6 & -15 & 9 \end{bmatrix}$$

$$\mathbf{B} = 3(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}})(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}})^{\mathrm{T}} + 3(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}})(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}})^{\mathrm{T}} = \frac{1}{6}\begin{bmatrix} 4 & -10 & 6 \\ -10 & 25 & -15 \\ 6 & -15 & 9 \end{bmatrix}$$

e)

$$\mathbf{W}^{-1}\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 1 & 4 \end{bmatrix}$$

$$\det\left|\mathbf{W}^{-1}\mathbf{B} - \lambda\mathbf{I}_2\right| = \det\left|\begin{bmatrix} 5-\lambda & 0 \\ 1 & 4-\lambda \end{bmatrix}\right| = (5-\lambda)(4-\lambda)$$

By setting this determinant to zero we obtain the eigenvalues of $\mathbf{W}^{-1}\mathbf{B}$ as the roots of $(5-\lambda)(4-\lambda)$, that is $\lambda_1 = 5$ and $\lambda_2 = 4$. Therefore, the optimal value of Fisher's discriminant is

$$\max_{\mathbf{a}\in\mathbb{R}^2}\left\{\frac{\mathbf{a}^{\mathrm{T}}\mathbf{B}\mathbf{a}}{\mathbf{a}^{\mathrm{T}}\mathbf{W}\mathbf{a}}\right\} = 5\,.$$