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

Exercise 10

Friday, January 11, 2019

Problem 1. *(k-means Clustering)*

The set $\Phi = \{\mathbf{x}_i \mid i = 1, \dots, 6\}$ contains 2-dimensional data which belongs to 2 clusters $\mathcal{C} \in \{1, 2\}$, with

$$\mathbf{x}_1 = \begin{pmatrix} 7 \\ 0 \end{pmatrix}, \ \mathbf{x}_2 = \begin{pmatrix} 7 \\ 3 \end{pmatrix}, \ \mathbf{x}_3 = \begin{pmatrix} 9 \\ 1 \end{pmatrix}, \ \mathbf{x}_4 = \begin{pmatrix} 9 \\ 5 \end{pmatrix}, \ \mathbf{x}_5 = \begin{pmatrix} 3 \\ 7 \end{pmatrix}, \ \mathbf{x}_6 = \begin{pmatrix} 12 \\ 3 \end{pmatrix}.$$

The k-means clustering algorithm is used to cluster the samples for Φ.

a) At a certain iteration, $\mathbf{x}_1$ and $\mathbf{x}_3$ are the center of cluster 1 and cluster 2, respectively. Assign each data sample in Φ to the appropriate cluster.

b) Update the centers of the clusters according to the assignment in **(a)**.

c) Suppose that the Euclidian distance in the k-means clustering algorithm is replaced by the following distances

$$d_1(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_1 = \mid x_1 - y_1 \mid + \mid x_2 - y_2 \mid,$$
$$d_\infty(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_\infty = \max(\mid x_1 - y_1 \mid, \mid x_2 - y_2 \mid),$$

for any $\mathbf{x}, \mathbf{y} \in \Phi$, with $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$.

Assign the data samples in Φ to the appropriate cluster, assuming $\mathbf{x}_1$ and $\mathbf{x}_3$ are the centers of cluster 1 and cluster 2, respectively.

Problem 2. *(Maximum Likelihood Clustering)* Suppose that $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ are n samples from g populations, each with Gaussian distribution $N_p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$. The corresponding densities are:

$$f_k(\mathbf{u}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{u} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1}(\mathbf{u} - \boldsymbol{\mu}_k)\right\}, \mathbf{u} \in \mathbb{R}^p.kl = 1, \dots, g.$$

a) Define the cluster analysis problem as maximization of log-likelihood function and write down the respective optimization problem.

b) Given clustering of samples $S_1, \dots, S_g$, find ML-estimation of $\boldsymbol{\Sigma}$.

c) Show that if $\boldsymbol{\Sigma}$ is unknown, the ML-cluster analysis is equivalent to the following optimization problem:

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

where

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

d) If $\mathbf{\Sigma}$ is known, show that ML-cluster analysis is equivalent to the following optimization problem:

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

Problem 3. *(Discriminant Analysis)*

A training dataset consists of three-dimensional vectors belonging to two classes (also known as groups) denoted by the labels $y_i \in \{1, 2\}$. The dataset is given below.

Data	Label	Data	Label
$\mathbf{x}_1 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$	$y_1 = 1$	$\mathbf{x}_4 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$	$y_4 = 2$
$\mathbf{x}_2 = \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}$	$y_2 = 1$	$\mathbf{x}_5 = \begin{pmatrix} 0 \\ 2 \\ -1 \end{pmatrix}$	$y_5 = 2$
$\mathbf{x}_3 = \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}$	$y_3 = 1$	$\mathbf{x}_6 = \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}$	$y_6 = 2$

a) Find the centering matrices, namely $\mathbf{E}_1$ and $\mathbf{E}_2$.

b) Find the average of the dataset, namely $\bar{\mathbf{x}}$.

c) Find the averages over groups 1 and 2, namely $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$.

d) Find the matrix $\mathbf{B}$ corresponding to the sum of squares between groups.

Now consider a different dataset where the inverse of the matrix $\mathbf{W}$ corresponding to the sum of squares within groups, and the matrix $\mathbf{B}$ corresponding to the sum of squares between groups, are given by

$$\mathbf{W}^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 3 & 2 \\ 2 & -2 \end{bmatrix}.$$

e) In Fisher discriminant analysis the maximum value of $\frac{\mathbf{a}^{\mathrm{T}}\mathbf{B}\mathbf{a}}{\mathbf{a}^{\mathrm{T}}\mathbf{W}\mathbf{a}}$ over all $\mathbf{a} \in \mathbb{R}^2$ is needed. Calculate the value of

$$\max_{\mathbf{a}\in\mathbb{R}^2} \frac{\mathbf{a}^{\mathrm{T}}\mathbf{B}\mathbf{a}}{\mathbf{a}^{\mathrm{T}}\mathbf{W}\mathbf{a}}.$$

Hint: there is no need for calculating the vector $\mathbf{a}$ that maximizes $\frac{\mathbf{a}^{\mathrm{T}}\mathbf{B}\mathbf{a}}{\mathbf{a}^{\mathrm{T}}\mathbf{W}\mathbf{a}}$.