**高等模糊控制**

**Advanced Fuzzy Logic Control**

**Program Report II**

電機所控制組

授課老師： 李祖聖 老師

姓 名： 廖聲宇(N26121143)

邱承浩(N2612)

目錄

[一、 題目概述 3](#_Toc153728112)

[二、 模糊分類設計步驟 5](#_Toc153728113)

[1. Fuzzy C-Means Clustering Algorithm 5](#_Toc153728114)

[2. Fuzzy K-Nearest Neighbor Approach 5](#_Toc153728115)

[3. Nearest Neighbor Approach 6](#_Toc153728116)

[三、實驗結果 7](#_Toc153728117)

# 題目概述

Apply the fuzzy c-means clustering algorithm, fuzzy k-nearest neighbor approach, and nearest neighbor approach to the **Exasens Data Set**, which is referred to

**UCI Machine Learning Repository [**[**http://archive.ics.uci.edu/ml/index.php**](http://archive.ics.uci.edu/ml/index.php)**]. Irvine, CA: University of California, Center for Machine Learning and Intelligent Systems.** (<https://archive.ics.uci.edu/ml/datasets/Exasens>)

一張含有 桌 的圖片

自動產生的描述

**Data Set Information:**

The Exasens dataset includes demographic information on 4 groups of saliva samples (COPD-Asthma-Infected-HC) collected in the frame of a joint research project, Exasens ([Web Link]), at the Research Center Borstel, BioMaterialBank Nord (Borstel, Germany). The sampling procedure of the patient materials was approved by the local ethics committee of the University of Luebeck under the approval number AZ-16-167 and a written informed consent was obtained from all subjects. A permittivity biosensor, developed at IHP Microelectronics (Frankfurt Oder, Germany), was used for the dielectric characterization of the saliva samples for classification purposes ([Web Link]).

Definition of 4 sample groups included within the Exasens dataset:

(I) Outpatients and hospitalized patients with COPD without acute respiratory infection (COPD).

(II) Outpatients and hospitalized patients with asthma without acute respiratory infections (Asthma).

(III) Patients with respiratory infections, but without COPD or asthma (Infected).

(IV) Healthy controls without COPD, asthma, or any respiratory infection (HC).

**Attribute Information:**

1- Diagnosis (COPD-HC-Asthma-Infected); 2- ID; 3- Age; 4- Gender (1=male, 0=female);

5- Smoking Status (1=Non-smoker, 2=Ex-smoker, 3=Active-smoker);

6- Saliva Permittivity: a) Imaginary part (Min(Î”)=Absolute minimum value, Avg.(Î”)=Average), b) Real part (Min(Î”)=Absolute minimum value, Avg.(Î”)=Average)

**Citation Request:**

The authors acknowledge the Federal Ministry for Education and Research (BMBF) of Germany for funding this work. The authors thank the BioMaterialBank Nord (BMB Nord), popgen 2.0 network (P2N), and the German Center for Lung Research for the collection of saliva samples and the staff at IHP and FZ Borstel-Leibniz Lung Center for their precious support with this work. In case of using the introduced Exasens dataset please cite the following papers:

P. S. Zarrin, N. Roeckendorf, and C. Wenger. In-vitro Classification of Saliva Samples of COPD Patients and Healthy Controls Using Non-perceptron Machine Learning Tools. Annals of biomedical engineering, 2020.

**Exclude all the data with Missing Values.**

**Part I: Perform the feature (attribute) extraction, clustering, and classification to make a correct diagnostic, where testing dada is the training data.**

**Part II: Perform the feature (attribute) extraction, clustering, and classification. For classification problem, please perform the leave-one-out method.**

**Questions:**

**Do you have to use all the attributes?**

**Is it possible to find the best combination of attribute to make the highest correction classification rate?**

# 模糊分類設計步驟

## Fuzzy C-Means Clustering Algorithm

Step 1. 選擇群集數量和指數權重，

並初始化歸屬函數矩陣以及設定



Step 2. 透過計算每一群中心



Step 3. 透過計算新的歸屬函數矩陣



假如，則(2)式成立，其餘則設定

Step 4. 判斷遞迴終止條件，首先計算，假設，則設定。並 回到Step2，反之則停止迭代。其中為終止參數，G為導出矩陣之範數。

## Fuzzy K-Nearest Neighbor Approach

Step 1. 設定，其中為欲分群數量。

Step 2. 初始化，並計算第i筆資料到第j群中心點之距離。

其中為第i筆資料；為第j群中心點，計算方式如(2)式所示。

Step 3. 計算歸屬函數



其中為Step2所計算的第i筆資料到第j群中心點之距離

其中為第i群中的第j群歸屬函數，如下二式所示。





Step 4. 比較不同歸類的歸屬函數值，值較大者即屬於該群集。

## Nearest Neighbor Approach

Step 1. 設定，其中為欲分群數量。

Step 2. 初始化，並計算第i筆資料到第j群中心點之距離。

其中為第i筆資料；為第j群中心點，計算方式如(2)式所示。

Step 3. 計算歸屬函數



Step 4. 比較不同歸類的歸屬函數值，值較大者即屬於該群集。

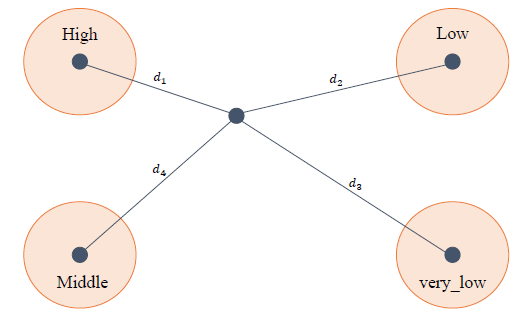
1. 實驗結果
   1. **Fuzzy C-Means Clustering Algorithm**
   2. **Fuzzy K-Nearest Neighbor Approach**
   3. **Nearest Neighbor Approach**

**Part I. Testing dada is the training data.測試資料等於訓練資料集。**

**A.**

首先要找到Testing dada中每個群集的中心，再將Testing data帶回訓練結果，透過訓練資料與各個中心的距離來判斷資料的歸屬群集，最後將結果比對Testing data並計算成功率。

各群集中心我們透過將各群集中的五筆資料(STG、SCG、STR、LPR、PEG)分別取平均，再把輸入的資料與各群集中的五筆資料取距離，進離資料較近的群集中心則為該筆輸入資料的群集。

****

其中

取min()後即可得最接近的群集。最後我們找到各群的中心，並得到成功率為75.36%。

表格 1 各群集中心

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | STG | SCG | STR | LPR | PEG |
| High | 0.4217 | 0.4232 | 0.5017 | 0.5013 | 0.7725 |
| Low | 0.3211 | 0.3370 | 0.4307 | 0.4973 | 0.2376 |
| Middle | 0.4000 | 0.3679 | 0.5068 | 0.3428 | 0.5424 |
| very\_low | 0.3058 | 0.1981 | 0.3663 | 0.3587 | 0.0908 |

B.

接著我們加入一個權重進行訓練，分別對五筆資料進行加權，並比較不同權重下的結果。

當權重如下表時，此時的成功率為20.93%。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | STG | SCG | STR | LPR | PEG |
| 權 重 | 1 | 0 | 0 | 0 | 0 |

當權重如下表時，此時的成功率為28.68%。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | STG | SCG | STR | LPR | PEG |
| 權 重 | 0 | 1 | 0 | 0 | 0 |

當權重如下表時，此時的成功率為27.91%。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | STG | SCG | STR | LPR | PEG |
| 權 重 | 0 | 0 | 1 | 0 | 0 |

當權重如下表時，此時的成功率為38.76%。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | STG | SCG | STR | LPR | PEG |
| 權 重 | 0 | 0 | 0 | 1 | 0 |

當權重如下表時，此時的成功率為78.29%。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | STG | SCG | STR | LPR | PEG |
| 權 重 | 0 | 0 | 0 | 0 | 1 |

由上述測試可知五筆資料中PEG資料對群集歸類的影響較大，此時的成功率比為加入權重時上升了2.93%。

**Part II. The leave-one-out method.將一筆資料取出做為測試資料。**

**A.**

首先隨機選擇資料作為leave-one-out數據，將剩餘的資料如前面的步驟所述進行訓練，最後把選擇出來的資料回去測試。

我們進行了10次leave-one-out的測試，最後得到的成功率為80%。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 是否正確 | T | T | T | T | T | T | F | T | F | T |

B.

我們同樣加入權重進行訓練，並分別進行10次leave-one-out的測試。

當權重如下表進行10次測試時，此時的成功率為20%。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | STG | | SCG | | STR | | LPR | | PEG | |
| 權 重 | 1 | | 0 | | 0 | | 0 | | 0 | |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 是否正確 | F | F | F | F | F | F | F | T | F | T |

當權重如下表進行10次測試時，此時的成功率為30%。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | STG | | SCG | | STR | | LPR | | PEG | |
| 權 重 | 0 | | 1 | | 0 | | 0 | | 0 | |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 是否正確 | F | F | F | F | F | T | T | F | T | F |

當權重如下表進行10次測試時，此時的成功率為20%。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | STG | | SCG | | STR | | LPR | | PEG | |
| 權 重 | 0 | | 0 | | 1 | | 0 | | 0 | |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 是否正確 | F | F | F | F | F | T | F | F | T | F |

當權重如下表進行10次測試時，此時的成功率為30%。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | STG | | SCG | | STR | | LPR | | PEG | |
| 權 重 | 0 | | 0 | | 0 | | 1 | | 0 | |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 是否正確 | T | T | F | T | F | F | F | F | F | F |

當權重如下表進行10次測試時，此時的成功率為60%。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | STG | | SCG | | STR | | LPR | | PEG | |
| 權 重 | 0 | | 0 | | 0 | | 0 | | 1 | |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 是否正確 | F | T | F | F | T | T | T | T | T | F |

由上述測試結果依然可以看出PEG資料對群集歸類的影響較大，PEG權重為1時的成功率相對其他權重高了30%~40%。

# 問題與討論