SathyaNarayanan Srinivasan

Sxs142031@utdallas.edu

Abstract

Perform WEKA related clustering and solves Bayes Network questions

Bayes network

Machine Learning Assignment – 6

Table of Contents

[Part - 1 2](#_Toc437290275)

[Part - 2 6](#_Toc437290276)

[Requirements 6](#_Toc437290277)

[Compiling and Running Instructions 7](#_Toc437290278)

[EXAMPLES 7](#_Toc437290279)

[**Output** 8](#_Toc437290280)

# Part - 1

1. **Clustering without PCA**
2. K-means with k=2: SSE =12598.193775711037

**Clustered Instances**

0 33 ( 87%) ALL(T&B)

1. 5 ( 13%)(AML)
2. **Confusion matrix for KMeans k=2**

|  |  |  |  |
| --- | --- | --- | --- |
|  | | PREDICTED | |
| AML | ALL-(T&B) |
| ACTUAL | AML | 4 | 7 |
| ALL-(T&B) | 1 | 26 |

1. **K-means with k=3:**

SSE= 11186.928018360504

**Clustered Instances**

0 20 ( 53%)( ALL B-cell)

1 3 ( 8%)(ALL-T-Cell)

1. 15 ( 39%)(AML)
2. **Confusion matrix for KMeans k=3**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | PREDICTED | | |
| AML | ALL-(T cell) | ALL-(B cell) |
| ACTUAL | AML | 1 | 3 | 7 |
| ALL-(T cell) | 3 | 0 | 5 |
| ALL-(B cell) | 11 | 0 | 8 |

1. **Hierarchical Clustering with K=2:**

**Clustered Instances**

0 37 ( 97%)(ALL-T& B cell)

1. 1 ( 3%)(AML)
2. **Confusion Matrix for Hierarchical Clustering K=2**

|  |  |  |  |
| --- | --- | --- | --- |
|  | | PREDICTED | |
| AML | ALL-(T&B) |
| ACTUAL | AML | 0 | 11 |
| ALL-(T&B) | 1 | 26 |

1. **Hierarchical Clustering with K=3:**

**Clustered Instances**

0 36 ( 95%)(ALL B-cell)

1 1 ( 3%)(ALL T-cell)

1. 1 ( 3%)(AML)
2. **Confusion matrix for Hierarchical Clustering K=3**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | PREDICTED | | |
| AML | ALL-(T cell) | ALL-(B cell) |
| ACTUAL | AML | 0 | 0 | 11 |
| ALL-(T cell) | 0 | 0 | 8 |
| ALL-(B cell) | 1 | 1 | 17 |

1. **Clustering after PCA**
2. **K-means with k=2:**

SSE = 56.3089363060435

**Clustered Instances**

0 28 ( 74%)(ALL T& B cell)

1. 10 ( 26%)(AML)
2. **Confusion matrix for KMeans k=2**

|  |  |  |  |
| --- | --- | --- | --- |
|  | | PREDICTED | |
| AML | ALL-(T&B) |
| ACTUAL | AML | 4 | 7 |
| ALL-(T&B) | 6 | 20 |

1. **K-means with k=3:**

SSE= 55.22769565994147

**Clustered Instances**

1. 20 ( 53%)(ALL B CELL)
2. 7 ( 18%)(ALL T CELL)
3. 11 ( 29%)(AML)
4. **Confusion matrix for KMeans k=3**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | PREDICTED | | |
| AML | ALL-(T cell) | ALL-(B cell) |
| ACTUAL | AML | 3 | 2 | 6 |
| ALL-(T cell) | 4 | 1 | 3 |
| ALL-(B cell) | 4 | 4 | 11 |

1. **Hierarchical Clustering with K=2:**

**Clustered Instances**

1. 37 ( 97%)(ALL-T& B cell)
2. 1 ( 3%)(AML)
3. **Confusion Matrix for Hierarchical Clustering K=2**

|  |  |  |  |
| --- | --- | --- | --- |
|  | | PREDICTED | |
| AML | ALL-(T&B) |
| ACTUAL | AML | 0 | 11 |
| ALL-(T&B) | 1 | 26 |

1. **Hierarchical Clustering with K=3**:

**Clustered Instances**

0 36 ( 95%)(ALL B-cell)

1. 1 ( 3%)(ALL T-cell)
2. 1 ( 3%)(AML)
3. **Confusion matrix for Hierarchical Clustering K=3**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | PREDICTED | | |
| AML | ALL-(T cell) | ALL-(B cell) |
| ACTUAL | AML | 1 | 0 | 10 |
| ALL-(T cell) | 0 | 0 | 8 |
| ALL-(B cell) | 0 | 1 | 18 |

1. **Classification**

|  |  |  |
| --- | --- | --- |
| Algorithm | Parameters | %Training accuracy |
| Random Forest | Max depth, No.of features, number of trees, seeds | 100 |
| Boosting | Classifier, number of iterations, seed | 5.2632 |
| J48 | Confidence factor, min no.of obj, number of folds, seed, subtree raising | 44.7368 |
| Bagging | Bag size percent, number of iterations,classifier ,seed | 55.2632 |

|  |  |  |
| --- | --- | --- |
| Algorithm | Parameters | %Test Accuracy |
| Random Forest | Max depth, No.of features, number of trees, seeds | 5.2632 |
| Boosting | Classifier, number of iterations,seed | 2.6316 |
| J48 | Confidence factor, min no.of obj, number of folds, seed, subtree raising | 0 |
| Bagging | Bag size percent, number of iterations,classifier ,seed | 2.6316 |

# Part – 2

Page

Page 2

# 

Page 3

# 

Page 4

# 

Page 5

# 

Page 6

# 