Report

Net ID: sxi140530

Name: Srikant Iyengar

Part 1:

1) **Clustering without PCA:**

**K-Means**

1. K-means with k=2 : SSE = 11597.541950968196

Seeds: 17

**Confusion Matrix:**

|  |  |  |
| --- | --- | --- |
| **Predicted/ Actual** | **0** | **1** |
| **Cluster 0 (AML)** | 14 | 1 |
| **Cluster 1 (ALL-B)** | 6 | 17 |

1. K-means with k=3 : SSE = 11515.397940192994

Seeds: 3

**Confusion Matrix:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Predicted/ Actual** | **0** | **1** | **2** |
| **Cluster 0 (AML)** | 10 | 0 | 1 |
| **Cluster 1 (ALL-T)** | 2 | 2 | 4 |
| **Cluster 2 (ALL-B)** | 8 | 1 | 10 |

**Hierarchical Clustering**

1. Hierarchical with k=2 :

Distance Function: ChebyshevDistance

**Confusion Matrix:**

|  |  |  |
| --- | --- | --- |
| **Predicted/ Actual** | **0** | **1** |
| **Cluster 0 (ALL-B)** | 30 | 0 |
| **Cluster 1 (ALL-T)** | 7 | 1 |

1. Hierarchical with k=3 :

Distance Function: ChebyshevDistance

**Confusion Matrix:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Predicted/ Actual** | **0** | **1** | **2** |
| **Cluster 0 (ALL-B)** | 18 | 0 | 1 |
| **Cluster 1 (ALL-T)** | 6 | 1 | 1 |
| **Cluster 2 (AML)** | 10 | 0 | 1 |

2) **Clustering after PCA:**

PCA done on 50 attributes.

Clustering results are as follows:

**K-Means**

1. K-means with k=2 : SSE = 85.48833160555867

Seeds: 17

**Confusion Matrix:**

|  |  |  |
| --- | --- | --- |
| **Predicted/ Actual** | **0** | **1** |
| **Cluster 0 (ALL-B)** | 18 | 5 |
| **Cluster 1 (ALL-B)** | 6 | 9 |

1. K-means with k=3 : SSE = 76.31386016335361

Seeds: 3

**Confusion Matrix:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Predicted/ Actual** | **0** | **1** | **2** |
| **Cluster 0 (AML)** | 2 | 1 | 8 |
| **Cluster 1 (ALL-T)** | 1 | 2 | 5 |
| **Cluster 2 (ALL-B)** | 2 | 4 | 13 |

**Hierarchical Clustering**

1. Hierarchical with k=2 :

Distance Function: ChebyshevDistance

**Confusion Matrix:**

|  |  |  |
| --- | --- | --- |
| **Predicted/ Actual** | **0** | **1** |
| **Cluster 0 (ALL-B)** | 27 | 0 |
| **Cluster 1 (AML)** | 10 | 1 |

1. Hierarchical with k=3 :

Distance Function: ChebyshevDistance

Link type : Complete

**Confusion Matrix:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Predicted/ Actual** | **0** | **1** | **2** |
| **Cluster 0 (ALL-B)** | 7 | 4 | 8 |
| **Cluster 1 (ALL-T)** | 2 | 3 | 3 |
| **Cluster 2 (AML)** | 2 | 1 | 8 |

3) **Classification**

**Training Data Accuracy**

|  |  |  |
| --- | --- | --- |
| Algorithm | Parameters | % Training Accuracy |
| Boosting | No of Iterations = 10,  Seed = 1 | 100 |
| Bagging | Bag size % = 100,  Seed = 1 | 100 |
| Naïve Bayes | Usekernelestimator = False,  Usesuperviseddiscretization = False, debug = False | 100 |
| Random Forest | Num trees = 100, Seed = 1 , max depth = 0 | 100 |
| J48 algorithm | numOfLeaves = 3; sizeOfTree = 5 | 100 |

**Test Data Accuracy**

|  |  |  |
| --- | --- | --- |
| Algorithm | Parameters | % Training Accuracy |
| Boosting | No of Iterations = 10,  Seed = 1 | 60.52 |
| Bagging | Bag size % = 100,  Seed = 1 | 57.8947 |
| Naïve Bayes | Usekernelestimator = False,  Usesuperviseddiscretization = False, debug = False | 60.52 |
| Random Forest | Num trees = 100, Seed = 1 , max depth = 0 | 50 |
| J48 algorithm | numOfLeaves = 3; sizeOfTree = 5 | 55.26 |