**Instructions:**

**There are a total of 5 (five) multi-part questions, with point values noted for each question.**

**Please show your calculations, or the details of your program(s), for each problem. Your program(s) should be commented so that each step is clearly explained.**

**Combine all of your answers/files into a single zipped file and post the zipped file to CANVAS.**

**Problems #1 and #2**

**Using an “Addiction” dataset, a researcher has prepared the following table of patient counts:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ethnicity** | **Age Category** | **Alcohol** | **Cocaine** | **Heroin** | **Row Total** |
| **Black** | **Old** | **30** | **48** | **17** | **95** |
|  | **Young** | **25** | **72** | **13** | **110** |
| **Hispanic** | **Old** | **7** | **0** | **5** | **12** |
|  | **Young** | **8** | **7** | **19** | **34** |
| **White** | **Old** | **60** | **2** | **17** | **79** |
|  | **Young** | **26** | **10** | **34** | **70** |
| **Column Total** |  | **156** | **139** | **105** | **400** |

**Use the table above and Excel to classify patient addiction type (alcohol, cocaine, heroin) using Ethnicity and Age Category:**

1. **Construct a classification and regression tree (CART) (two levels only).**  (20 Points)
2. **Construct a C4.5 decision tree (two levels only).**  (20 Points)

**Problem #3**

**Use R/python to cluster (Algorithm=K-means; K=2) the seven (7) already normalized points in the accompanying table and answer a and b below:** (20 points)

|  |  |  |  |
| --- | --- | --- | --- |
|  | X | Y | Z |
| **a** | **1** | **1** | **1** |
| **b** | **5** | **3** | **4** |
| **c** | **4** | **4** | **5** |
| **d** | **4** | **3** | **4** |
| **e** | **1** | **2** | **1** |
| **f** | **4** | **4** | **4** |
| **g** | **2** | **1** | **2** |

1. **What are the members of each cluster?**
2. **What are the coordinates for the cluster centers?**

**Problem #4**

**Use R/python and the above table (problem #3) to cluster (Algorithm=hierarchical; two clusters) the seven (7) already normalized points in the accompanying table and answer a and b below:** (20 points)

1. **What are the members of each cluster?**
2. **What are the coordinates for the cluster centers?**

**Problem #5**

**Using data in the table below, construct a Neural Network with one Output Layer (z) and one Hidden Layer (A and B).** (20 points)

1. **Calculate the predicted outcome if the inputs to the input nodes are (x=1, Node 1=.4, Node 2=.7 Node 3= .7 and Node 4=.2).**
2. **Adjust the weight if the actual output is 0.8500**

|  |  |  |
| --- | --- | --- |
| **From** | **To** | **Weight** |
| X | A | 0.5 |
| Node 1 | A | 0.6 |
| Node 2 | A | 0.8 |
| Node 3 | A | 0.6 |
| Node 4 | A | 0.2 |
| x | B | 0.7 |
| Node 1 | B | 0.9 |
| Node 2 | B | 0.8 |
| Node 3 | B | 0.4 |
| Node 4 | B | 0.2 |
| xx | z | 0.5 |
| A | z | 0.9 |
| B | z | 0.9 |