**Unless it is specified, you may use Word, Excel, Power Point and R to answer the questions in this exam. There are a total of 4 (four) multi-part questions, with point values noted for each question.**

**Please show your calculations, or the details of your program(s), for each problem. The R 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.**

**Problem #1: (10 points)**

**Using data in the table below, construct a Neural Network with one Output Layer (z) and one Hidden Layer (A and B). Calculate the predicted outcome if the inputs to the input nodes are (Node 1=.4, Node 2=.6 Node 3=.6 and Node 4=.2).**

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

**Problem #2: (30 points)**

**Cluster NJ zip codes into 3 clusters using the NJ income data in CANVAS, R and the following methods.**

1. **Hierarchical clustering and median income**
2. **K-means, population and median income**

**Problem #3: (30 points)**

* **Load the Titanic dataset from CANVAS**
* **Convert the categorical variables to 0,1 indicators**
* **Store every fifth record in a “test” dataset starting with the first record**
* **Store the rest in the “training” dataset**
* **Use ANN with 5 hidden nodes to classify passengers (survival=1 vs. 0).**
* **Measure the performance of the model against the test data.**

**Problem #4: (30 points)**

* **Load the breast cancer dataset ( wisc\_bc\_ContinuousVar .csv), with continuous variables, from CANVAS.**
* **Remove all the rows with any missing values**
* **Store every fifth record in a “test” dataset starting with the first record**
* **Store the rest in the “training” dataset**
* **Use Random Forest to classify the breast cancer cells.**
* **Score the test dataset and the model performance.**