CISC/CMPE452/CISC874/COGS 400

Assignment **2**

Implement a Backpropagation Network

Farhana Zulkernine

15 marks

General Instructions for Code and Submission (for all assignments)

1. You can use any programming language (preferred python, C, C++, Java, or Matlab)

2. Make **one zip file named as** Asg2_studentID.

3. **Upload zip file** to the OnQ site. Multiple uploads are allowed. Only the most recent version will be kept.

Assignment 2 : The Data Set

- The data set you are going to use for this assignment is the dataset for classifying 6 different glass types.

- There are 9 features, excluding the ID of the instance, that define the glass type listed in the last column as the Glass_Type. The 9 features are: refractive index, Sodium, Magnesium, Aluminium, Silicon, Potassium, Calcium, Barium, and Iron.

- The different glass types are:
1 = building_windows_float_processed,
2 = building_windows_non_float_processed,
3 = vehicle_windows_float_processed,
5 = containers, 6 = tableware, 7 = headlamps

- There are 214 instances given in the CSV file.

- You need to do necessary preprocessing of the data to use it for training, validating and testing your BP ANN based on the engineering design approaches we discussed in the class.

Assignment 2 : Tasks
(15 marks)

1. Design a Back-propagation ANN based on the given data set. Marks will be given for accurate coding and algorithm (**2 marks**), clarity, comments (**1 mark**) and ability to execute your program (**2 marks**). You must provide all necessary components and instructions to execute your code otherwise you will lose points.

2. Note down and ***explain your design choices in a text file***. You must submit the following in the text file.

   - Initial weights, node output function used, leaning rate, termination criteria with proper explanations for the choice **(2 marks)**

   - Number of layers and nodes used and why **(1 mark)**

   - Momentum parameter value **(0 for none, 0.5 mark)**

   - Regularization approach used **(0 for none, 0.5 mark)**

   - Data preprocessing **(processing + explanation: 1 mark)**

   - Percentage of data used for training, validation and testing and how the split was done. Remember to review the design principles of ANN. **(simple split and explanation: 1 mark, correct approach to use proper data distribution and explanation: 2 marks)**

Assignment 2 : Tasks (cont…)
(15 marks)

3. Generate performance statistics of your ANN in terms of

   - Precision (**0.5 mark**), recall (**0.5 mark**) and a confusion matrix (**1 mark**)

   - Final weight vectors (**0.5 mark**)

   - A text/CSV file listing given class label and your program generated class label. Use the function from your last assignment to create the output file. (**0.5 mark**)

   Tip: You can use a tool like Matlab or Weka to construct your ANN model first and have an idea about what kind of structure (number of layers, number of nodes in each layer) will result in a higher accuracy for the given problem. (No mark for this)

Submit

- The program code with comments, executable code (if applicable). (#1 last slide)

- A text file with all design choices and performance statistics.

- A text file with the output – given class and predicted class labels against each row of data values.

- Combine all the above in one zip file and upload that on OnQ.

- Missing submission will be marked as 0.

- The penalty for late submission past the due date will be -1 mark per day.