Quiz # 6. AMS 580

Name:_____SBU ID:_____

Dear all, you have the entire lecture time to do this quiz. It is open book however you must do the work entirely on your own. Please turn on your video and mute your audio.

Please submit your Quiz in three documents: (1). The Rmd file, (2). The output file (word or pdf), and (3). The testing data with predicted responses added (csv file) – to the SBU Brightspace. If you have difficulty submitting to Brightspace, you may then email the solutions (with the subject of "Quiz 6, AMS 580", and in one email with the three attachments), to your TA Ian at: weihao.wang@stonybrook.edu

Neural Network with the GreatUnknown Data – Classification Task

The **GreatUnknown.csv** data contain 12 predictors and one binary response variable **y** (= 0 or 1), which is the true class label**.**

For this dataset, **sensitivity** is defined as a case labeled 1 being classified to label 1, while **specificity** is defined as a case labeled 0 being classified to label 0.

1. For the entire dataset, please perform the data cleaning as instructed before; namely, delete observations with missing value(s). Please report how many cases (namely, data points) are left after this step. Then please use the random seed 123 to divide the cleaned data into 75% training and 25% testing.

2. Please first build the best classifier to predict **y** using the training data and the Perceptron model with (i) no hidden layer, (ii) the default loss function of "sse", and (iii) the default activation function of "logistic". Please plot the perceptron model obtained using the training data. Please compute the Confusion matrix and report the sensitivity, specificity, and the overall accuracy using the testing data.

3. Next we will build the best classifier to predict **y** using the training data and the Perceptron model with (i) no hidden layer, (ii) the loss function of "ce" (namely, cross-entropy, or the negative log likelihood), and (iii) the default activation function of "logistic". Please plot the perceptron model obtained using the training data. Please compute the Confusion matrix and report the sensitivity, specificity, and the overall accuracy using the testing data.

4. Now we shall build the best classifier to predict **y** using the training data and the Logistic Regression model. Please report the fitted logistic regression model obtained using the training data – and compare to the Perceptron models obtained in the plots of Question 2 and Question 3. Which Perceptron model better resembles the logistic regression model, and why? Please compute the Confusion matrix and report the sensitivity, specificity, and the overall accuracy

using the testing data.

5. Now we shall build the best classifier to predict **y** using the training data and the Perceptron model with (i) one hidden layer with 3 neurons, (ii) the default loss function of "sse", and (iii) the default activation function of "logistic". Please plot the perceptron model obtained using the training data. Please compute the Confusion matrix and report the sensitivity, specificity, and the overall accuracy using the testing data. Please compare the performance in test data to that of Question 2.

6. Next we shall build the best classifier to predict **y** using the training data and the Perceptron model with (i) one hidden layer with 3 neurons, (ii) the loss function of "ce" (namely, cross-entropy, or the negative log likelihood), and (iii) the default activation function of "logistic". Please plot the perceptron model obtained using the training data. Please compute the Confusion matrix and report the sensitivity, specificity, and the overall accuracy using the testing data. Please compare the performance in test data to that of Question 3.

7. Which neural network model provides the best overall accuracy among Parts **6** and **7** above? For this best model only, please add the predicted label for every test data point, and then output the test data as a .csv file.