Practice questions

The dataset contains various measurements (i.e. size, center, etc) from thousands of bacterium under microscope. The last column with non-zero values indicate the bacterium are interesting enough for further study. Otherwise, those bacterium are not interesting candidates for further study. Convert this dependent variable to binary values. Normalize predictors first using Z-score.

Write a MatLab (or a programming language of your choice) program to use Autoencoder to predict the class of bacteria records. You can select your own Autoencoder architecture (i.e. use any number of layers) to find the best classification result. **Remember** the quality measurement should be obtained from testing data, not training data.

Answer the following questions and put your answers in **a WORD document**:

1. Show your Autoencoder architecture.

2. Describe your Autoencoder in details (i.e How many layers do you have? How

many neurons in each layer? What is the dimensions of the code you used in

classification? Did you change anything else in the Autoencoder to increase the

prediction quality?)

3. Show your confusion matrix. What is the overall accuracy? What are Precision,

Recall, and F-score of each class prediction?

4. Create an ROC curve plot for each class prediction.