**BENG 420/520: Homework #4 Neural Networks**

**Due: Wednesday, 4/15/2020, by 11:59pm**

*Instructions: You can either work by yourself or find another fellow in the class to work on it. You should submit this assignment as a Word or PDF document through Blackboard. Please also submit your commented Matlab code.*

1. The objective of this assignment is to practice using Artificial Neural Networks in real life biomedical problems. You are asked to implement a simplified version of the cancer classification method described in Khan et al’s paper (ANN gene nature 2001.pdf). You will need to use functions in the Matlab Neural Network toolbox.
2. h4\_data.mat contains the training data and the test data. ‘geneexpTrain’ and ‘geneexpTest’ are the feature matrices of the training data and test data respectively. The data has been mapped to **ten** dimensions as described by the paper.

* Note that I have eliminated those 5 test instances from the 25 instances in the original test dataset that are healthy samples (i.e. not belonging to any class).

1. H4\_ANN.m is the template file for you to program your ANN classifier.
2. **Build** a one-layer **linear** neural network as proposed by the paper. **Train** the network using the 63 training instances. You can decide how many instances to use for training/validation/test in training the neural network. If you decide to use all for training and none for validation/testing, that is also OK. **Justify** why you make that choice.
3. **Examine** the training accuracy and test accuracy by predicting class labels of the training instances and test instances using your trained ANN classifier. **Plot** the confusion matrices for both cases. Include snapshots of the confusion matrix plots in your report. **Explain** why your result is not as perfect as those reported in the paper. **Discuss** your observation on the performance of the trained classifier by examining the training accuracy and test accuracy (e.g. overfitting, underfitting, model complexity, etc.)
4. Apply (and revise if needed) the k-NN algorithm you implemented in Homework #1 to this dataset. Program in h4\_kNN.m. You will need to **determine** the optimal ‘k’ in the way similar to what you did previously. **Plot** the test accuracy as a function of ‘k’. **Include** a snapshot of the plot in your report.
5. **Compare** the best test accuracy of k-NN to the test accuracy of your ANN classifier and **describe** what you observe. **Discuss** why k-NN results in that kind of performance. For instance, do factors such as dataset size and dimensionality affect k-NN for this particular problem? What does the result imply?
6. **Submit** a report, h4\_ANN.m, h4\_kNN.m, and other supplementary Matlab code you did.