ECE426/526 Homework on Machine Learning.

**In your submission, please list all team members. You should have a word/pdf document to directly answer questions. Submit your Matlab code either as a .m or .mix file.**

A Step-by-Step guide is found in Machine-Learning-workflow-ebook.pdf, which is online (also under Files/Lecture material). Go over the document

Download *MATLAB code for the heart sounds classification application.*

After the download, you should have a folder having many files. The main file is the HeartSoundClassification.mlx. You need to understand the code step-by-step and see how classification is done and make changes to show your understanding and answer questions.

In your submitted work, please answer the following questions:

* Plot the power spectral densities of two additional normal and abnormal PCG’s (the document has an example for each). Do all the three abnormal PCG’s show a clear spike near 0.2 radian?
* How many \*.wav files are there in the data/training directory (need to use Matlab to count)? How many training files have abnormal and normal PCG’s?
* How many \*.wav files are there in the data/validation directory (need to use Matlab to count)? How many \*.wav validation files have abnormal and normal PCG’s?
* What is the size of the variable feature\_table? Report the means and standard deviations of the 28 features for abnormal and normal PCG’s. By comparing the means, which features are likely most/least important for the classification.
* Run five classifiers using classificationLearner from at least three different groups (all models in Decision Trees are considered to be in the same group). You should have at least one classifier having an accuracy below 75% and one above 92%. Present the confusion matrices for each. Briefly explain the principle each classifer uses (in a few sentences).
* For your best model, reduce the false negative to less than 1% by modifying the cost function.
* In the Perform Feature Selection using Neighborhood Component Analysis section, what are the three most and least important features? Support your response with your own data.

In the above, you do not change the feature\_table. For the following tasks, remove the 12 least important features in the original feature\_table, do the following:

* Rerun the same five classifiers using classificationLearner as you did before. Compare the confusion matrices with the previous runs. How do the accuracies compare?
* What are the three most and least important features in the new feature\_table?