HW9 (for graduate students only)

DUE: Sunday 6/4/17 by 11:59 pm

You are hired by the National Science Foundation to work on the problem of classifying glitches. Glitches are non-Gaussian disturbances in the gravitational-wave data of the Advanced Laser Interferometer Gravitational-wave Observatory (aLIGO).

Glitches are typically represented as spectrogram, a time-frequency representation where the x-axis represents the duration of the glitch and the y-axis shows the frequency content of the glitch. The colors indicate the "loudness" of the glitch in the aLIGO detector. For this assignment, however, glitches are represented as vectors and they cannot be easily visualized.

Different environmental and instrumental mechanisms will produce glitches of different shape and morphology. Here, we consider four major classes Blip, Whistle, Koi fish and Wandering line. Your task is to build a classification model which could help scientists understand the underlying mechanisms for creating such glitches so that they are removed from the detector.

You are given the file *train_data_label.csv*. Each row of this matrix represents one data sample, where the last column is an integer representing the corresponding label, i.e.: 1 is Blip, 2 is Whistle, 3 is Koi fish and 4 is Wandering line.

You are also given the file *test_data.csv*. Each row of this matrix is one data sample, and the corresponding label is missing.

Use your favorite algorithm from the ones covered in class to build a classifier using *train_data_label.csv* in order to predict the label for *test_data.csv*.

Submit the labels as a single csv file:

1. No header.
2. Each label is on a new line, without a comma. Therefore your .csv file will only consist of one column; each row is the label for the corresponding row in test_data.csv.
3. You are expected to achieve over 98% test accuracy. Every 1% less will result in a deduction of 2 points.
4. Named as your netid.csv.