The goal of this course is to provide an introduction in the area of machine learning using real data. Real data are acquired by each student speaking the ten English digits 0 through 9 into the microphone of a laptop. Programs written in the Matlab programming language convert the microphone speech into numerical features that are then processed to classify the speech sounds into their digit values. Matlab programs are provided by the instructor to illustrate the theory.

The course covers template matching and neural networks for classifying spoken digits. The performance is assessed through a confusion matrix that estimates the probability of classification error. By the end of the course, each student will have composed real-time Matlab programs that classify spoken digits.

A student-specific project tailored to the student’s experience and interests will be assigned early in the course. At the last class meeting, each student will give an Powerpoint presentation and submit a report that describe their approach.

Course topics:

1. Overview
   a. Speech signal
   b. Signal features
      i. Cepstral coefficients (Shift invariant)
      ii. Spectrograms (threshold aligned)
   c. Training (Template) and Test sets.
   d. Real-time speech classification
   e. Goal: Minimize classification error probability
   f. Approaches
      i. Template matching
      ii. ANN for linear classification
      iii. ANN with hidden layers for non-linear

2. Template matching
   a. Correlation processor
   b. Similarity index
   c. Nearest neighbor in feature space
d. Confusion Matrix
e. Probability of error
f. kth-Nearest neighbor

3. Neural network (NN) using linear classification
   a. Orthogonal feature vector
   b. Gradient descent of loss function
   c. Back-propagation for weight updates
   d. Training
   e. Testing
   f. Confusion Matrix and Probability of error
   g. Real-time bit classification

4. Speech classification
   a. Supervised learning using label data
   b. Non-orthogonal feature vectors
   c. Target values of the output
   d. NN forms weights

5. NN with hidden layers using non-linear classification
   a. Speech acquisition using laptop microphone
   b. Spectrogram templates for NN training
   c. Variable-width time windows to normalize spectrogram time scale
   d. Increasing template separation
   e. Test set of recorded digits
   f. Confusion Matrix and Probability of error
   g. Real-time digit classification

6. Speech feature and NN alternatives
   a. Merging cepstral coefficients and spectral features
   b. Four-output NN for classifying binary digit representations
   c. Binary features.
   d. Summary of current speech recognition NNs
      i. Feature pooling to reduce feature vector size
      ii. Convolutional NN

Grading:
- Homework – 20%
- Midterm exam – 30%
- Final report and presentation – 50%

Course reading list:
There is no introductory textbook that covers the material in this course. Instead of a textbook, class notes and reprints of relevant material from the internet will be posted on the class web page. Links to useful on-line videos will be provided. Searching and learning from such on-line material is a valuable technique for continued learning.
**Required preparation:**

1. Basic knowledge of any programming language for understanding how to implement an algorithm.
2. Become familiar with Matlab basics by using on-line tutorial *MATLAB Onramp*
   [https://matlabacademy.mathworks.com/](https://matlabacademy.mathworks.com/)
3. Familiarity with vector calculations (inner product) in matrix algebra.