

## Assignment 1: Intro to Machine Learning

### Total points 20

Consider the data matrix ([data.csv](#)) containing 20 datapoints each for two different classes (class 1 and 2) with two features f1 (column 1) and f2 (column2). The class labels are given in column 3 in the file.

- a. Construct frequency tables with 5 bins of equal width for each feature given the class. Draw the frequency curve/histogram of the features for each class.
- b. Find estimation of  $P(10 < f1 < 20 \mid c1)$  and  $P(10 < f1 < 20 \mid c2)$  from the frequency tables.
- c. Classification using single features: Suppose we use maximum likelihood classifier, that is if
$$P(a < f1 < b \mid c1) > P(a < f1 < b \mid c2)$$
 then we classify the data point with f1 value within the range  $[a,b]$  as class c1, otherwise c2.
  - i. What threshold value on f1 would you use to classify the data using f1 alone? Using the frequency table find the optimal threshold t1. What is the optimum error rate?
  - ii. What threshold value on f2 would you use to classify the data using f2 alone? Using the frequency table find the optimal threshold t2. What is the optimum error rate?
- d. Classification using both features:
  - i. Plot the data (using a scatter plot. You may use matlab/matlibplot in python). If you use t1 and t2 threshold found, you will divide the space (feature space) into 4 regions. What will be the classification decision for data falling in each region? How would you resolve the conflicts?
  - ii. Draw a linear decision boundary by hand that you think classifies the data best. What would be the equation for the decision boundary?

### Submission Instruction:

1. For this assignment you should do all the calculations by hand and use matlab/octave/python to draw the figures. Put the code and figures in your doc and save it as a pdf.
2. You do not need to do prepare your submission using latex but will receive extra credits if you actually use latex.
3. Name your pdf file [yourid]\_assign1.pdf (use your actual id in place of '[yourid]').

**Submission deadline: Oct 6, 2021, by 11:59am.**