

## Project 5: Bayes Classifier for Iris Plant Classification and Breast Cancer Classification

### Each group should

- Submit one lab report per group to Canvas by May 31, 11:59pm.
- Include a thoughtful and reflective paragraph from each team member summarizing what has been learned.
- Include a section describing the work done by each member of the group.
- Attach your well-documented code in the appendix of the report.
- For this project, please don't use MATLAB's built-in functions from Statistics and Machine Learning Toolbox.

In this project, you will use Bayes classifier to perform two tasks – iris plant classification, and Breast cancer classification.

### 1. Data Sets

- i. The Iris flower data set consists of 50 samples from each of three species of Iris (Setosa, Virginica and Versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimeters.

You can find the Iris data here

<http://archive.ics.uci.edu/ml/datasets/Iris>

The data set is also available in MATLAB (*load fisheriris*)

- ii. The Breast Cancer Wisconsin (Diagnostic) data set consists of 30 features computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. A total of 569 instances were collected of which 357 were benign and 212 malignant.

You can find the data here

<http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/>

Use `wdbc.data`

### 2. Design and evaluation of the classifier

- i. Perform feature analysis and selection  
You can use all features or a subset of the features to perform classification. Show your analysis and justify your feature selection.
- ii. Train Bayes classifier for the two tasks  
For Iris plant classification, use “leave-one-out” procedure to select training and test data. For the Breast Cancer data set, use 80% for training and the remaining 20% for test.

## iii. Evaluate the classifier

Use confusion matrix and probability of error to report the performance for Iris plant classification. Determine TP, TN, FP, and FN for Breast Cancer classification assuming FP and FN are equally costly (i.e. use zero-one loss function). Plot the ROC curve.

## iv. Naïve Bayes classifier

Build Naïve Bayes classifier for the two tasks and compare its performance with the Bayes classifier.