

The goal of this program assignment is to let you experiment with classifiers. You can use Matlab, python, or C++. You are not allowed to use existing libraries/modules/toolboxes for classification/evaluation, but libraries for linear algebra operations are allowed.

Your program should include the following modules:

1. The classifier modules (one for each classifier type), with separate methods for training and testing. The training method takes training data and their labels as inputs. The output method takes testing data as input and returns both the predicted classes of the testing data and values of the discriminant functions. For this assignment, you need to implement the following classifiers:

- Bayesian classifier; use Gaussian pdfs with parameters given by maximum-likelihood estimation.
- Naïve-Bayes classifier
- One of the linear classifiers covered in the class (only to be applied to two-class datasets)

2. The evaluation module. This module uses the outputs of the classifier module for the testing data, as well as their ground-truth class labels, to do evaluation of the classification results. You should compute and display the confusion matrices, and the ROC curves (as well as AUCs) for two-class datasets. Use these values to compare the different classifiers.

In addition, in the main program, you need to split the dataset into training and testing subsets, with both sets having similar distributions. We are also going to introduce cross-validation in class shortly; this will allow you to evaluate the classifiers more systematically.

Go to the [UCI Machine Learning Repository](#) to get datasets for your experiments. Choose only datasets for classification and having numerical attributes. The different datasets have different data file formats so you need to carefully read the descriptions. Some datasets that have been used a lot in the literature include: [Iris](#), [Breast Cancer](#), [Glass](#), [Ionosphere](#), [Optical Recognition of Handwritten Digits](#), and [Wine](#). You can go through the list of hundreds of datasets to find others to play with. Select at least two two-class datasets and two datasets with more classes.

You need to do experiments with your classifiers. Try changing things and see what happens. Examples include using subsets of features or different numbers of training samples.

Write a report (limited to 10 pages, single-spaced) describing

- Methods you have implemented.
- Experiments you have done, and the results.
- Analysis - Are the results what you expect? Why?

The report (Word or PDF formats) is to be submitted electronically through e3. Also include the program listing as an appendix to your report (in the same file, not counted toward the page limit).

Late submission policy: 10% credit deduction for each day late; up to 7 days late accepted.