

# CS423 Data Mining: Assignment 2

Due date: 7 Nov 2017

## Instruction

In this assignment, you will implement Bayesian classifiers for classifying “biodegradable” data from UCI repository. There are two variations of Bayesian classifiers that you need to investigate. Assuming that your data is normally distributed (Gaussian) implement:

1. Gaussian classifier having full covariance matrix for each class.
2. Gaussian classifier having shared covariance matrix with naive assumption.

Measure the performance of the two classifiers by cross-validation technique. That is to partition the data into two sets: training set and testing set. Train your classifier on the training set and test the learned classifier on the testing set. Record the classification errors and repeat the process for 10 times. Finally, report your classification errors together with their standard deviations from the 10 repetitions.

## Submission

Send the report and the codes to `jakramate.b@cmu.ac.th`. Prefixing the subject of your email with “[423HW3]”.