

CS423 Data Mining: Assignment 3

CMU, Due date: 29 November 2017

Instruction

In this assignment, you will compare predictive performance of a generative classifier represented by Gaussian classifier and a discriminative classifier represented by logistic regression. You have to decide first what type of Gaussian classifiers should be used in the experiment *in order to guarantee a fair comparison*. That is, will it be Gaussian classifier with shared or separate covariance matrix ?, will it be Gaussian classifier with full or diagonal covariance ? You need to write your choice down in the report. Next you will train the logistic regression (using the code implemented in the lab session) as well as the Gaussian classifier of your choice (that you have already implemented in homework 2) on the following datasets

1. Australian
2. Banknote
3. Boston
4. Fertility
5. Pima

Describe your experiment protocol in details. Namely, explain how did you measure the classifiers performances in order to ensure unbiased comparison. Summarise the results in a table and discuss your results in depth.

Submission

Submit the codes to `jakramate.b@cmu.ac.th`, prefixing your email subject with [423HW3].