

NICTA - Statistical Machine Learning advanced course

Second assignment: Binary classification

The goal of this assignment is to compare empirically different classifiers on a real data set

1. propose a method to provide a fair comparison between different classification methods on empirical data using 10 fold-cross validation to tune hyperparameters.
2. Go on the web and download the SPAM data base at UCI:

`ftp://ftp.ics.uci.edu/pub/machine-learning-databases/spambase/`

LDA Write a Matlab routine implementing the linear discriminant analysis (LDA) algorithm. Apply it on the SPAM UCI data base.

k -NN Write a Matlab routine implementing the k -nearest neighbour classifier. Apply it on the SPAM UCI data base.

Parzen Write a Matlab routine implementing the naive Bayes classifier based on Parzen window estimate. Apply it on the SPAM UCI data base.

SVM Write (or down load) a Matlab routine implementing the SVM estimate. Apply it on the SPAM UCI data base.

3. Compare your results according to the error rates and computing time. Discuss and compare with available state-of-the-art results. Any suggestion to improve the results?

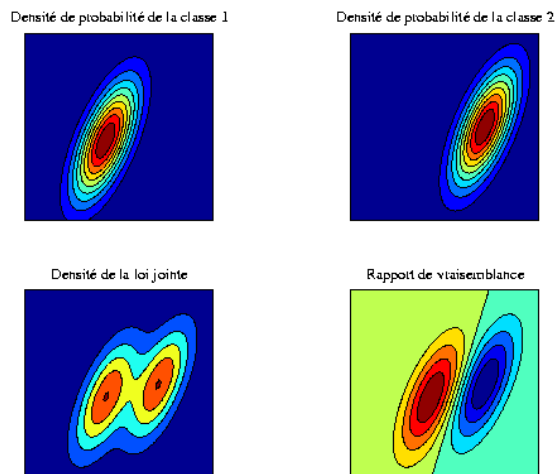


Figure 1: LDA discriminates optimally two gaussian class.