

Artificial Intelligence Final Project

Presentation by Raffaello Camoriano, MSc in Robotics Engineering
University of Genoa, Faculty of Engineering - A.A. 2011/2012



Introduction

Goals:

- Tackling a binary classification problem with Machine Learning techniques.
- Performing benchmark dataset analysis.
- Comparing different classifiers.

Dataset Overview

General description:

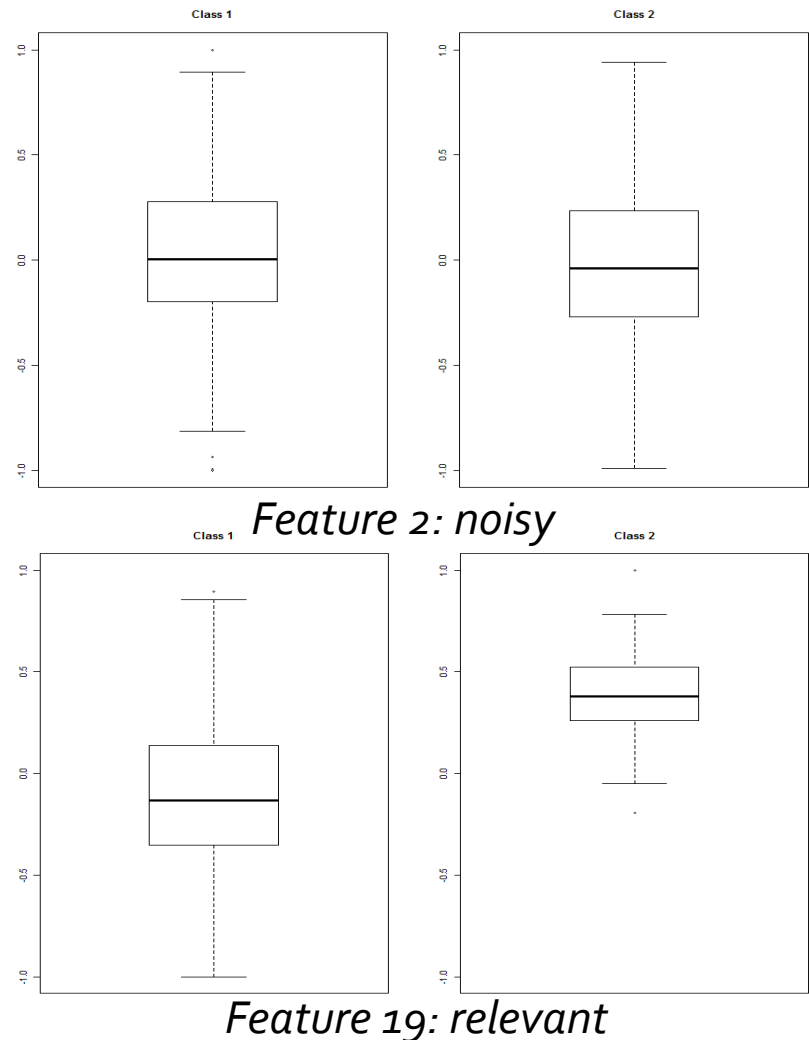
- 25 features, some of which are noisy.
- 121 samples
- No NULL values
- Few samples: curse of dimensionality

Feature Analysis (1)

- Identify relevant and irrelevant features

Qualitative tools:

- Boxplots
- Correlogram

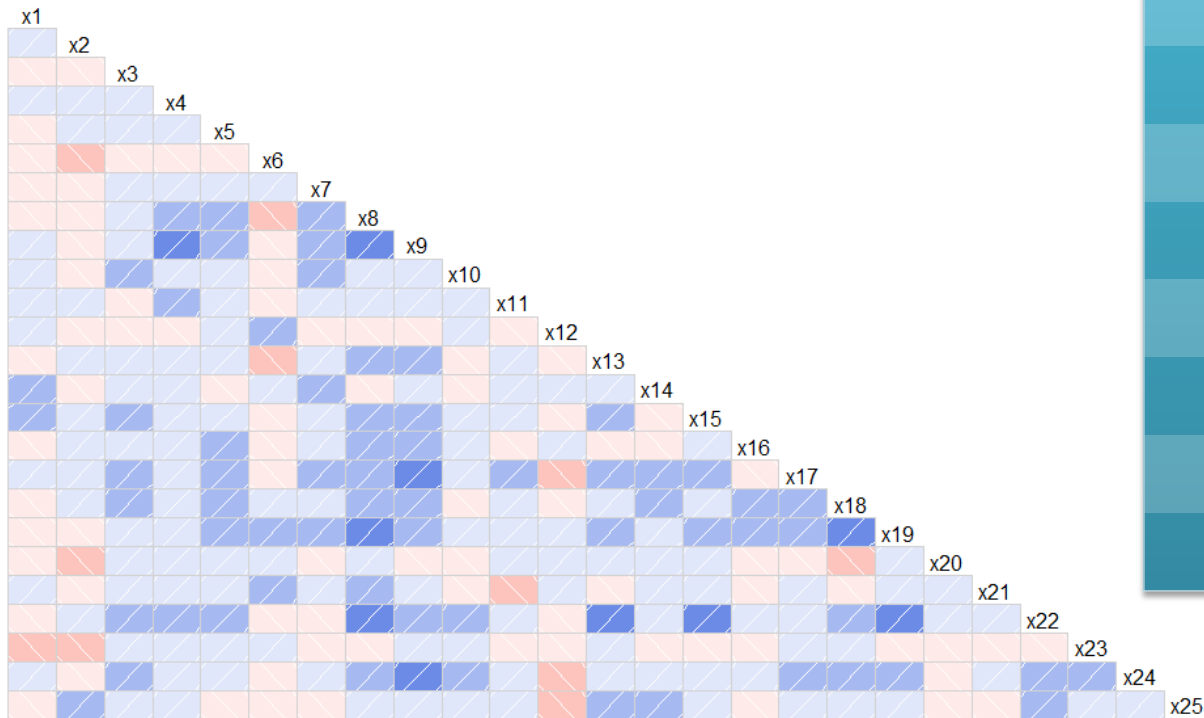


Feature Analysis (2)

■ Correlogram

Red: negative correlation

Blue: positive correlation



Positive

x4, x9

x8, x9

x8, x19

x8, x22

x9, x17

x9, x24

x13, x22

x15, x22

x18, x19

x19, x22

Negative

x1, x23

x2, x6

x2, x20

x2, x23

x6, x8

x6, x13

x11, x21

x12, x17

x12, x24

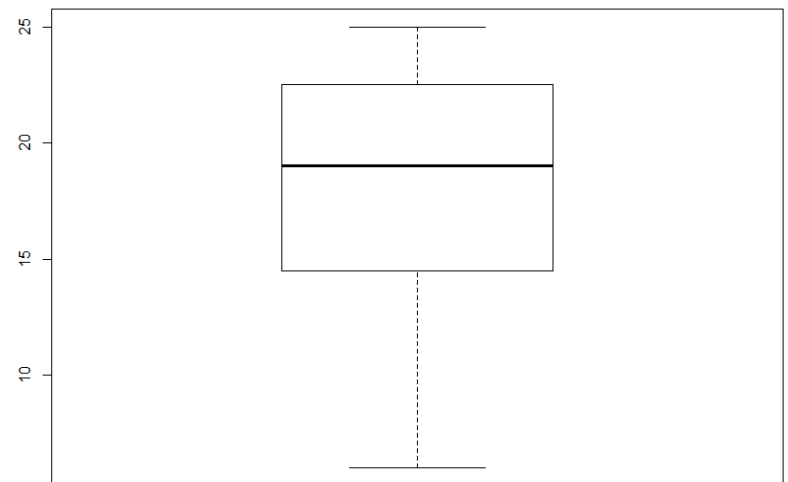
x12, x25

x18, x20

Feature Selection

Automated feature selection procedure:

- Introduction of 3 dummy features with normal distribution
 - Execution of the Relief algorithm
 - KNN-1 classifier is applied for error estimation
 - Only the best set of columns in 121 iterations is kept
-
- The median value of selected features is 18



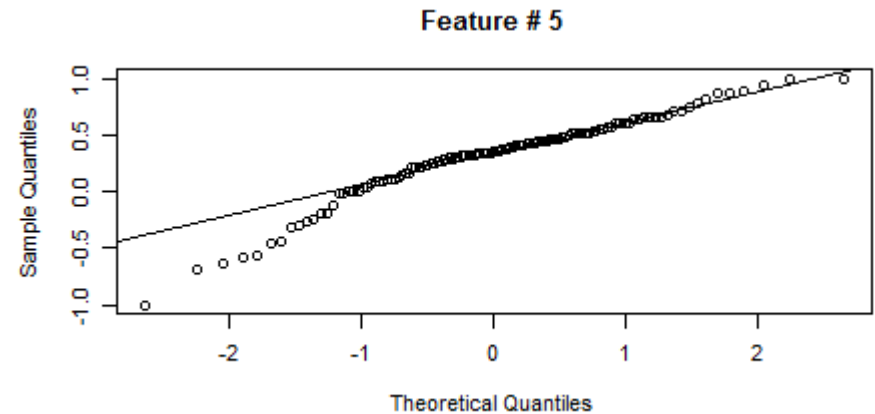
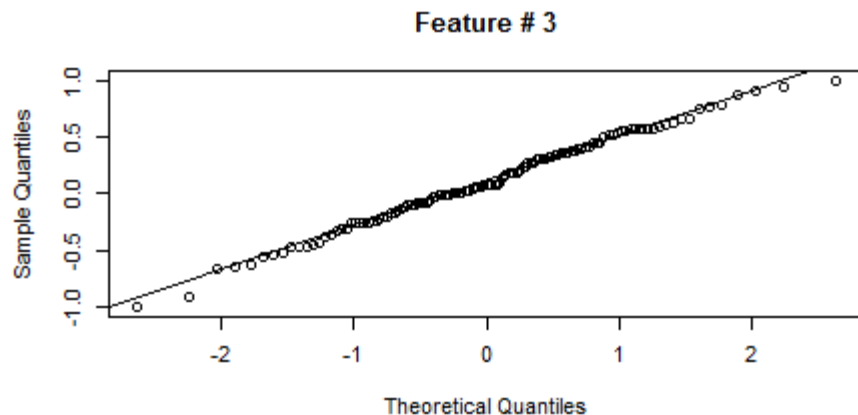
Normality Tests (1)

Two normality tests have been applied:

- Shapiro-Wilk
- Lilliefors (Kolmogorov-Smirnov)
- The results show that there are some features with a non-normal distribution.

Normality Tests (2)

- Q-Q Plots can help us to visualize the results of quantitative tests.



Feature Extraction

Issues:

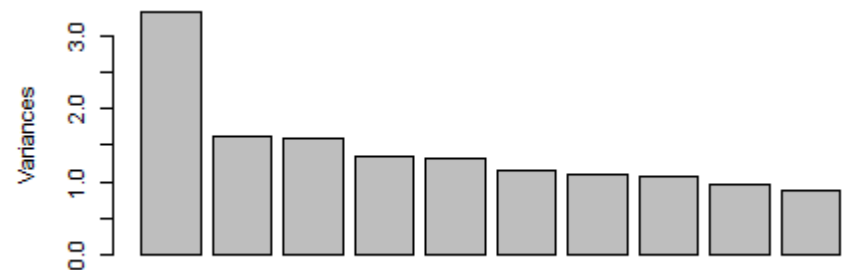
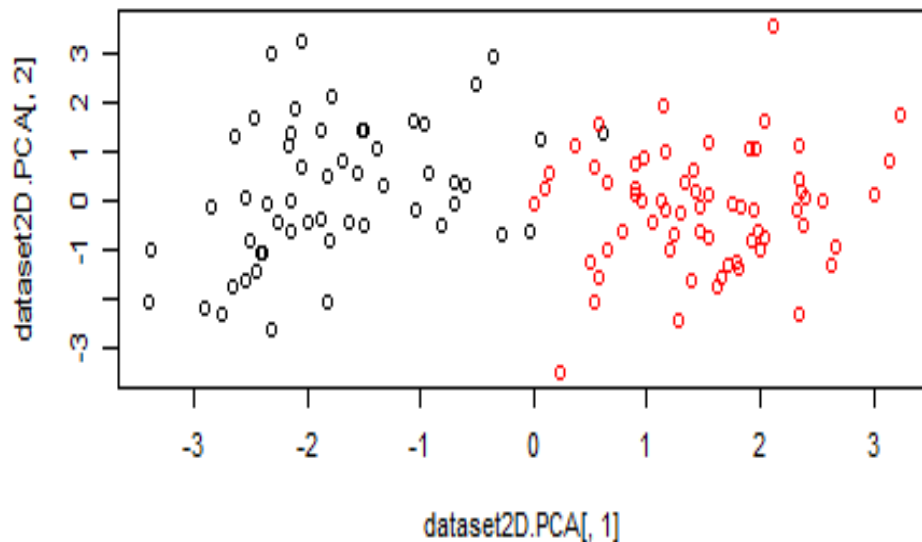
- Visualizing multidimensional data
- Reducing dimensionality of data

Possible solutions:

- **PCA** (Principal Component Analysis)
- **MDS** (Multidimensional Scaling)

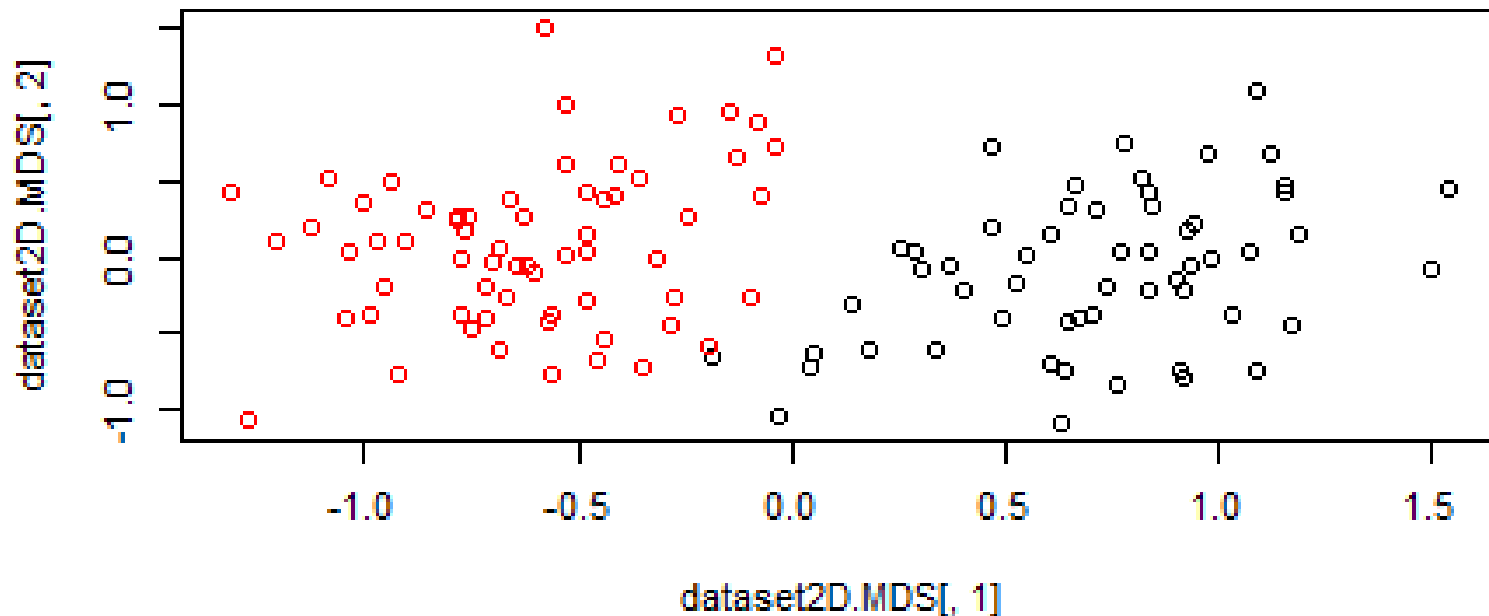
PCA

- Rotates data s.t. variance is maximised along each new axis. (a l.c. of the features is performed)
- Assumes normal distribution of features
- This is not true for all the features of dataset 18



MDS

- Less binding in terms of necessary assumptions
- Leads to a better result (clearly separable clusters)



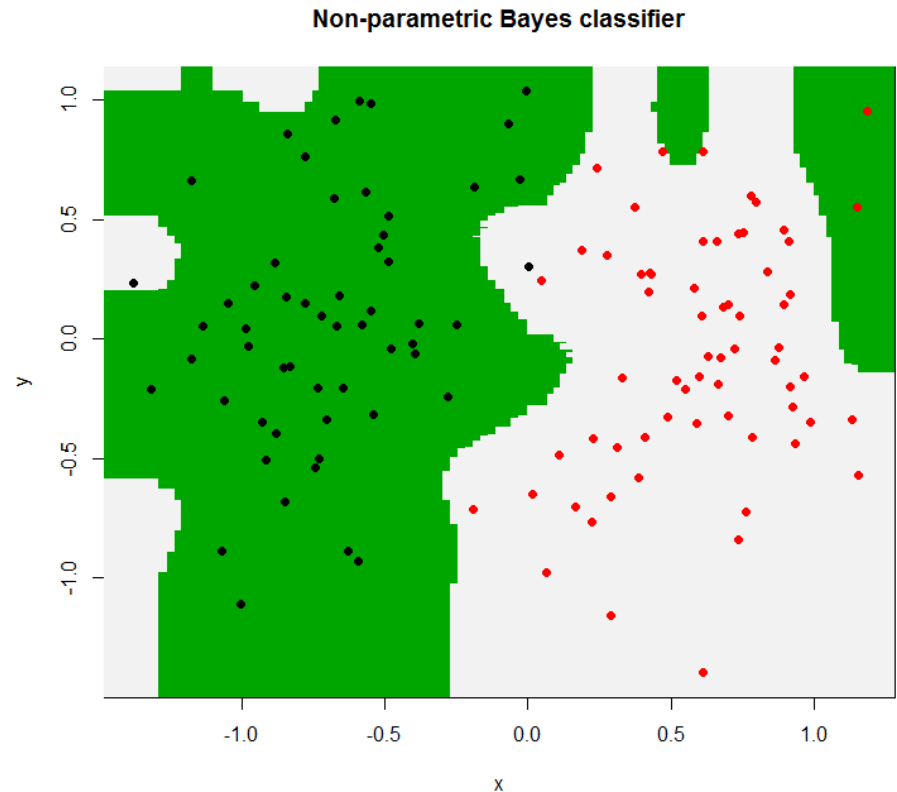
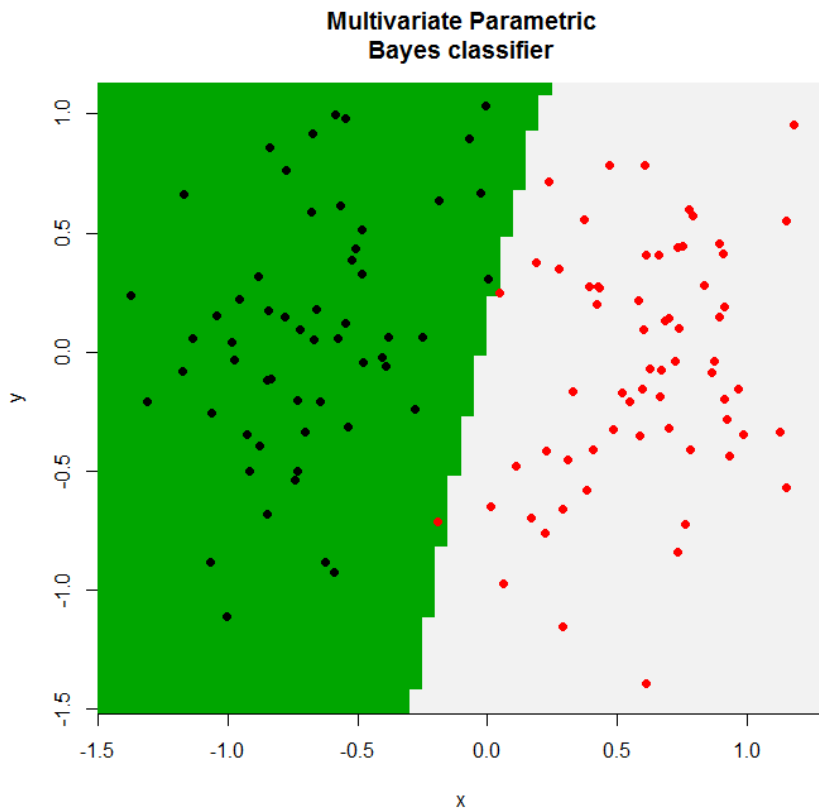
Classification

Tested classifiers:

- Naïve Bayes
- Multivariate Parametric Bayes
- Multivariate Non-parametric Bayes
- K Nearest Neighbours
- Linear Discriminant Analysis
- Single Hidden Layer Neural Network
- Support Vector Machine
- Decision Tree (rpart)

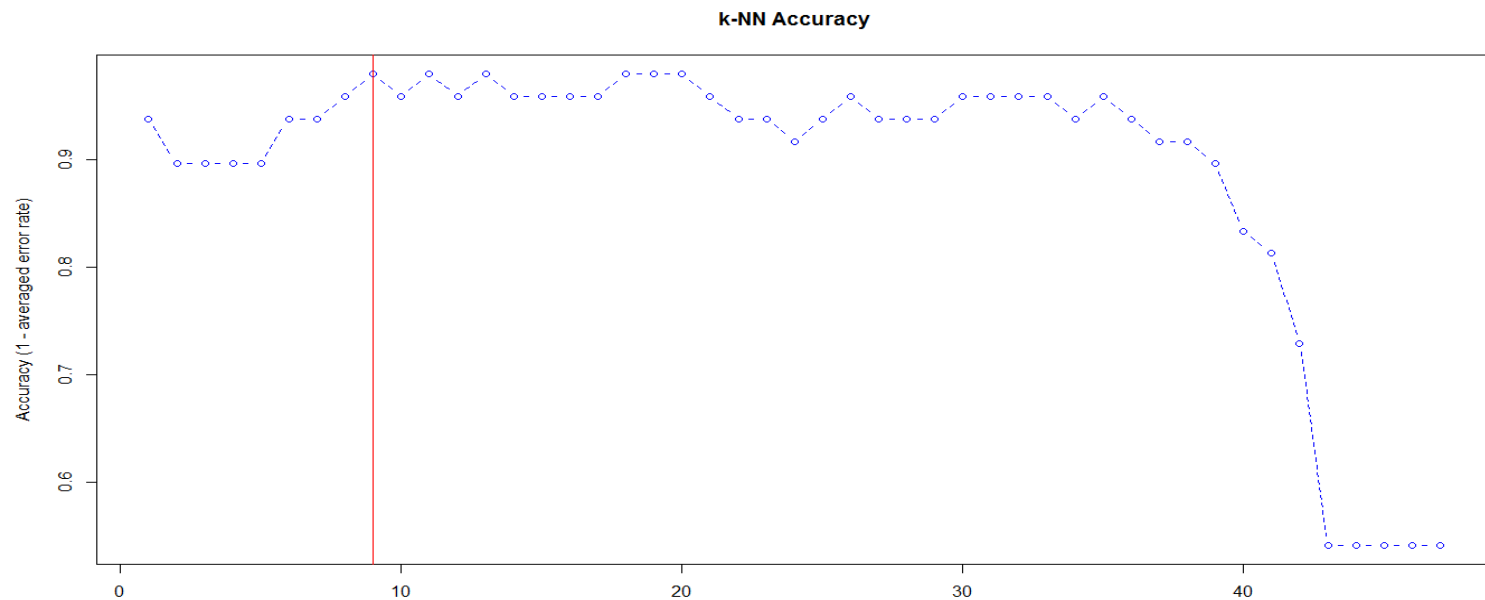
Multivariate Bayes

- Parametric and Non-parametric B. decision regions



Parameters tuning – Example: KNN

- Parameter K has been tuned
- The training set has been split in two equal parts, the first for optimization and the second for error estimation (Bootstrap)



Accuracy of KNN classification as a function of the K parameter

Final Results

Error rates of the classifiers (with 20 features):

- SVM has the best performance
- NB is more robust to changes of the number of features

