

Statistical Machine Learning Assignment

Agbatan Fiacre Luc KOUDERIN

African Institute for Mathematical Sciences (AIMS - Rwanda)

February 2, 2025



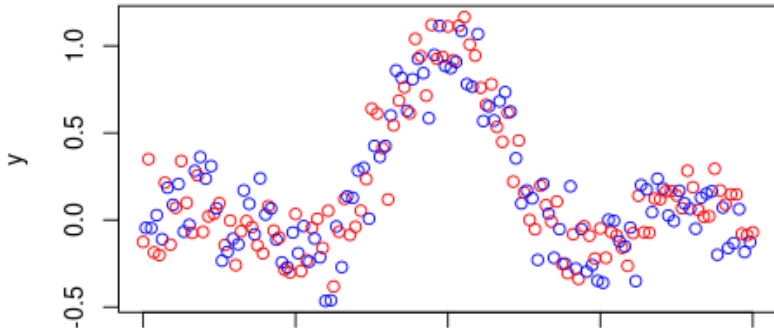
Outline

- Regression Model task
- Classification Model task

Regression task

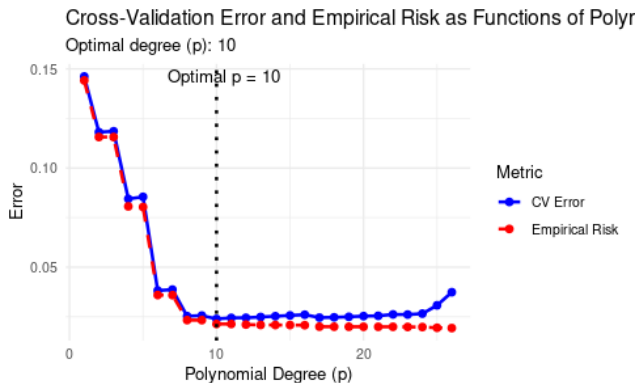
Data description

- 2 variables (one response variable, one covariate) with 225 observations
- Scatter plot of y versus x :



Regression task

Decision on the degree of polynomial

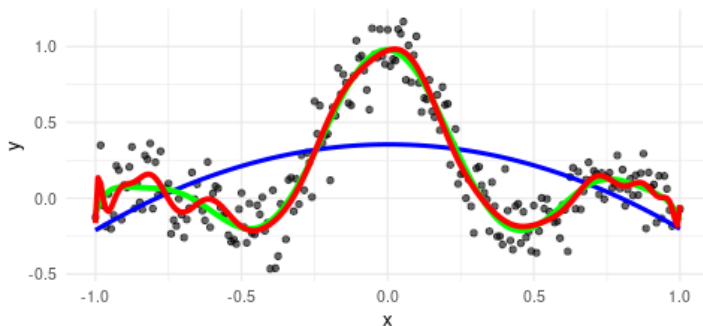


Regression task

Model fitted and comparison with another model

Simplest, Optimal, and Overly Complex Models

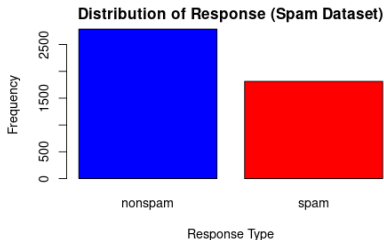
Model — Optimal Model — Overly Complex Model — Simplest Model



Classification task

Data description

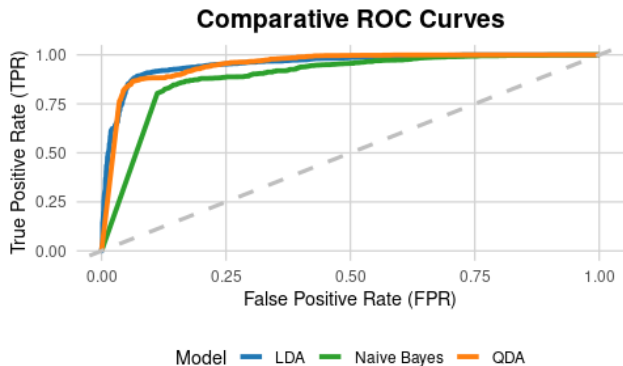
- Data **Spam** from the library **kernlab** in R
- 58 variables (one response variable, 57 covariates) with 4601 observations
- Goal : Use four machines : LDA, QDA, LFD and Naive Bayes
- Distribution of the response variable (2788 'non spam' and 1813



'spam') :

Classification task

Performance comparison on the training set



Classification task

Performance comparison on the test errors

