

# Assignment

Herp, Jürgen\*

December 7, 2017

The assignment covers a practical and a theoretical part. The practical part comprises the main corpus of the assignment.

**Keywords:** LDA,  $k$ -NN, GLM

- What are the models test error rates?
- Which model offers the best prediction?
- What are the relevant observed variables to consider?

## 1 Q-1: Boston Housing

The goal is to predict, as good as possible, whether a given suburb in Boston, USA, has a crime rate above or below the crime median.

- Explore the `Boston` dataset contained in the `MASS` library in R<sup>1</sup> or see the link for non R users<sup>2</sup>. What is the dataset comprised of?
- Fit classification models. Explore logistic regression, LDA, and  $k$ -NN models for that purpose. Consider also subsets of the observed variables. Include relevant parts of your scripts to complement your description of the models under consideration.
- Describe your findings. Include the following questions when elaborating on your finding:

---

\*herp@mmmi.sdu.dk

<sup>1</sup><https://cran.r-project.org/web/packages/MASS/index.html>

<sup>2</sup><http://lib.stat.cmu.edu/datasets/boston>

## 2 Q-2: Students Performance

Consider the following: Based on the final grade in data science classes, students were discriminated based on the time invested into their studies each week, and their grade point average. Assume further, that a discriminative logistic model proposes  $\beta_0 = -6$ , and  $\beta = [0.05, 1]^\top$ , for scoring an A (on the ECTS grade scale) in the finals. Let  $\mathbf{x} = [X_1, X_2]^\top$  be the model's variable, with  $X_1$ , the time invested by a student, and  $X_2$ , the student's grade point average.

- Write the explicit form of the model and estimate the probability that a student who studies for 40 hours a week, and has a grade point average of C<sup>+</sup> (equals 3.5 in numeric value), scores an A in the finals.
- How many hours a week would the student in part (a) need to spend to achieve a 80% chance of scoring an A?