

Assignment 2: Regression and classification

1. The dataset associated to this assignment was downloaded from www.hemnet.se on 2020-10-18. The data contains information about selling prices of villas in Landvetter that were sold in the previous 12 months.
 - a. Find a linear regression model that relates the living area to the selling price. If you did any data cleaning step(s), describe what you did and explain why.
 - b. What are the values of the slope and intercept of the regression line?
 - c. Use this model to predict the selling prices of houses which have living area 100 m², 150 m² and 200 m².
 - d. Draw a residual plot.
2. In this question, you will use the Iris data set ("from sklearn.datasets import load_iris").
 - a. Use a confusion matrix to evaluate the use of logistic regression to classify the iris data set.
 - b. Use k-nearest neighbours to classify the iris data set with some different values for k, and with uniform and distance-based weights. What will happen when k grows larger for the different cases? Why?
 - c. Compare the classification models for the iris data set that are generated by k-nearest neighbours (for the different settings from question b) and by logistic regression. Calculate confusion matrices for these models and discuss the performance of the various models.

Submitting work

Please append all python code written to your report PDF and submit everything as one PDF file. Give the names of the people submitting the work and state how many hours each person spent working on the assignment, on the first page of the report.

Remember, we check for plagiarism, and we are obliged to report suspected cases.

Deadline: Tuesday 31 January 2023 at 23:59.

Self-check

Is all the required information on the front page? Have you answered all questions to the best of your ability? Anything else you can easily check? (details, terminology, arguments, clearly stated answers etc.?)

It is important to:

- i. Present clear arguments
- ii. Present the results in a pedagogical way

- Should it be table/plot? What kind of plot? Is everything clear and easy to understand?
- iii. Show understanding of the topics
- iv. Give correct solutions.
- v. Make sure that the code is well commented.
 - Important parts of the code should be included in the running text and the full code uploaded to Canvas.