Data Science and Machine Learning in Python

Assignment 5 – Weather Prediction

Stephan Weyers

Fachhochschule Dortmund
University of Applied Sciences and Arts

Assignment 5

Fachhochschule Dortmund
University of Applied Sciences and Arts

Provided files

- W05_train.txt (training dataset)
- W05_test.txt (test dataset)
- W05_Results.xlsx
- W05_Task.pdf (this document)

Instructions

- Work on this assignment with the team mates who have been assigned to you
- Follow the instructions on the next page
- Solve the tasks using Python and Jupyter Notebook
- Fill in the results into the provided Excel-file
- Evaluate the contributions of all team members also in the Excel file
- Submit both ipynb and xlsx files in ILIAS

Due date

- June 28th (23:59 German time)

Assignment 5 – Exercises

Fachhochschule Dortmund
University of Applied Sciences and Arts

- Import the training set "W05_train.txt“ and test set "W05_test.txt"
- The last column „relative_humidity_3pm“ is the target variable, the first 9 columns are input variables
- Fit a linear regression model to the training data using all input variables.
  Report the R^2-score of the training and test set
- Fit a linear regression model to the training data using the first 6 input variables.
  Report the R^2-score of the training and test set
- Fit a linear regression model to the training data using the first 7 input variables.
  Report the R^2-score of the training and test set
- Fit a linear regression model to the training data using only the 7th input variable „relative_humidity_9am“.
  Report the R^2-score of the training and test set
- Fit a ridge regression model to the training data using all input variables with alpha=10.
  Report the R^2-score of the training and test set
- Fit a ridge regression model to the training data using all input variables with alpha=100.
  Report the R^2-score of the training and test set as well as the value of the constant and all coefficients of the model
- Fit a lasso regression model to the training data using all input variables with alpha=10.
  Report the R^2-score of the training and test set as well as the value of the constant and all coefficients of the model