

Intelligent Data Analysis

Exam: Fire in the nature park (Project 2)

This project is part of the exam *Intelligent Data Analysis*. Each project assignment is to be resolved by a single student on his/her own. The student is supposed to present the solution as part of the oral exam. The student is required to present a printed version of the Python code together with diagrams, tables, etc. that summarize the results. The specific way of how the project is presented is up to the student's choice.

Problem setting

The administration of the nature park *Montesinho* in north-east Portugal wants to predict wild fires based on weather data of the Fire-Weather-Index (FWI). The aim is to recognize the affected area and consequently the intensity of the imminent wild fire as early as possible in order to be able to adequately assess the danger caused by the fire. To this aim, data from 517 wild fires have been collected. The features are summarized below.

- X (X-coordinate of the fire in the park: 1 to 9)
- Y (Y-coordinate of the fire in the park: 2 to 9)
- month (month: „jan“ to „dec“)
- day (day: „mon“ bis „sun“)
- FFMC (FFMC index of the FWI system: 18.7 to 96.2)
- DMC (DMC index of the FWI system: 1.1 bis 291.3)
- DC (DC index of the system: 7.9 bis 860.6)
- ISI (ISI index of the FWI systems: 0.0 to 56.1)
- temp (temperature in degrees Celsius: 2.2 to 33.3)
- RH (relative humidity in %: 15 to 100)
- wind (wind velocity in km/h : 0.4 to 9.4)
- rain (rainfall in mm/m^2 : 0.0 to 6.4)
- area (forest area that has been burnt in hectare: 0.09 to 1090.84)

You have been asked to develop a model that predicts the burnt forest area as accurately as possible from the given data.

Exercise

Load the data into Python and preprocess them appropriately; perform an adequate normalization of the features. For example, the label *area* is distributed very non-uniformly such that a transformation such as $area' = \log(1 + area)$ is appropriate. Identify and implement a suitable learning method in Python. Train and evaluate the model. Propose a trivial baseline model with which you can compare your models performance. Apply a reasonable evaluation method. Provide a short documentation and motivation of each of your steps.