Aprenentatge Automàtic per a Xarxes (ML4Net)

Seminar 1 - Problem statement

April 2, 2025

Abstract

In this seminar, you will implement a linear regression algorithm that predicts the performance of Wi-Fi deployments. For that, you will use a reference dataset (`dataset_Seminar1.txt`) that contains the simulated Wi-Fi networks along with different parameters. The main purpose of this seminar is twofold: (1) process, visualize, and understand the target dataset, (2) train and evaluate a linear regression algorithm using the target dataset.

1 Part I: Data preparation & analysis

Create a module that processes the data. The main operations your module must do are:

- Loading the dataset from the .txt file and storing the data in a structured manner.
- Displaying statistics about each feature (mean, standard deviation).
- Plotting relevant figures to visualize how data is distributed.
- Processing the data (e.g., cleaning, normalization, categorical, etc.) and prepare it to be used by an ML model.
- Splitting the data into different partitions to be used for training (e.g., 80%) and testing (e.g., 20%).

2 Part II: Model implementation

Create a module that trains and evaluate a linear regression model. The main operations your module must do are:

- Fitting the data into an ML model (in this case, a multi-variable linear regression), i.e., estimating the coefficients of the model.
- Evaluating the performance of the model using a given error function.
- Plotting the results visually (e.g., training loss, test accuracy).