46765 - Machine Learning for Energy Systems
Assignment 1
Deadline: October 29th, 2025 (10:00pm)

Instructions: This assignment evaluates the topics covered in **Lectures 1 - 5** as well as programming and writing skills. It should be carried out in groups. Each group should provide a single submission, including the following:

- A concise project report (maximum 12 pages), detailing the mathematical models developed and presenting and analysing the main results. An appendix can be included (outside of page limit).
- A participation table accompanying the report, detailing the participation of each group member.
- A working and well-documented code in the programming language of your choice. Include also a README in your repository to explain the file structure you have and how to run the code.
- Additional relevant files and data.

All relevant files should be uploaded on DTU Learn. This assignment will count towards 33% of the final grade. The assessment will be based on the grading guide (rubric table) provided in the repository. Make sure to look at it when writing the report.

Objectives: The aim of Assignment 1 is to evaluate

- Your understanding of day-ahead and balancing electricity markets,
- Your ability to use real-world data as input,
- Your skills in data preprocessing, implementing linear and non-linear regression with regularization, and applying unsupervised learning techniques,
- Your understanding of classical supervised learning and decision-focused learning,
- Your critical analysis of the results generated.

Context: Envision yourself as the owner of a wind farm, aiming to participate in the day-ahead electricity market by leveraging historical data. For this assignment, you will work with:

- Actual wind power production data from the Snorrebakken wind farm in Bornholm: link
- Weather/climate data from nearby stations: link
- Day-ahead prices: link
- Balancing prices (up-regulation and down-regulation): link

You can download the required datasets from **EnergyDataDK** link. Please refer to the slides on DTU Learn for guidance.

Notes:

- Bornholm is located in the **DK2 market area**.
- Use **hourly resolution** for all data. If the data has finer granularity, aggregate it by taking the hourly average.
- Assume that weather observations represent forecasts (since no reliable forecast dataset is available).
- Time series data is incomplete due to issues such as database errors and communication failures:
  - Exclude periods with large gaps,
  - For smaller gaps, apply interpolation or mean imputation.
- Not all datasets need to be used.
- Assignments will be graded based on the rubric table uploaded to your repository

The goal is to **develop two models** for trading in the day-ahead market and to compare their performance:

1. A **classical regression model** for wind power prediction and bidding accordingly,
2. A **decision-focused learning model** for designing an optimal bidding strategy.

Model 1: Wind Power Prediction

In this approach, the task is to forecast wind power for the following day and use this forecast as the bidding quantity in the day-ahead market. As introduced in Lecture 2, regression techniques can be applied for prediction purposes. To begin, you should construct a dataset from the provided sources. Potential features may include: previous-day wind power, the 0.05, 0.5, and 0.95 quantiles of wind production from the past week, forecast date, and climate variables such as wind speed and temperature. These are only suggestions, and you are encouraged to select whichever features you believe improve predictive performance.

(a) **Evaluation methodology:** Explain how you partition the data into training, validation, and testing sets, and describe the role of each.

Note: Ensure that the same training, validation, and testing datasets are used consistently across all models developed in this assignment.

(b) **Dataset construction:** Describe the number and type of features selected, and explain both your quantitative feature selection method and the scaling or preprocessing applied.

(c) **Linear regression:** Train a linear regression model as a baseline. As discussed in Lecture 2, this problem can be approached in multiple ways, such as using the closed-form solution or optimization-based methods provided by Python libraries (e.g. Scikit-learn). Perform a quantitative comparison of these approaches in terms of both prediction accuracy and computational efficiency. Keep in mind that most modern machine learning packages implement gradient descent or one of its variants as the underlying solver for regression models.

(d) **Non-linear regression (feature modification):** Following Lecture 3, extend the linear regression model by applying feature transformations to capture non-linearities. Evaluate its performance and compare it with the baseline linear model.

(e) **Non-linear regression (weighted regression):** Extend the linear regression model to a nonlinear one by assigning weights to data points (locally weighted regression). Evaluate its performance and compare it with the baseline linear model.

(f) **Regularization:** As introduced in Lecture 4, apply regularization to mitigate overfitting. Implement both L1 (Lasso) and L2 (Ridge) regularization on linear and non-linear models. Tune the regularization parameter using the validation set, and assess how regularization affects model accuracy and feature coefficients. Compare results between L1 and L2 regularization.

Revenue Evaluation

Since the ultimate objective is profit maximization in the day-ahead market, calculate the revenue and cost resulting from bidding your forecasted production. Use historical day-ahead and balancing prices together with the predicted and actual wind power (from the test set) to quantify profits, as discussed in Lecture 2. Determine which model achieves the highest profit.

Unsupervised Learning

An advanced strategy for improving accuracy is to apply different regression models to different subsets of the data. To do this, employ a clustering technique from Lecture 5 to partition the data into groups. For each cluster, fit a local regression model from those studied earlier. Evaluate whether this method leads to improved predictions and discuss your findings.

Model 2: Decision-Focused Learning

Consider a decision-focused learning framework, as introduced in Lecture 5, to design an optimal bidding strategy.

- *Mandatory step*: Formulate the corresponding optimization problem used for training. Clearly define all elements of your formulation, including decision variables, parameters, the objective function, and any constraints.

- *Optional step*: Implement this formulation and compare the results of this model with your previous regression models. Summarize your findings and provide recommendations for further improving trading strategies. (This step provides an additional 10% points, but the total grade will not exceed 100%. You may exceed the page limit by up to one page.)